

JACKELINE TERTO DA SILVA SANTANA

**CARACTERIZAÇÃO GENÔMICA DO DNA REPETITIVO E DIVERSIDADE
GENÉTICA EM *Syagrus coronata* (Mart.) Becc.**

RECIFE - PE

2022

JACKELINE TERTO DA SILVA SANTANA

**CARACTERIZAÇÃO GENÔMICA DO DNA REPETITIVO E DIVERSIDADE
GENÉTICA EM *Syagrus coronata* (Mart.) Becc.**

Tese apresentada ao Programa de Pós-Graduação em Agronomia – Melhoramento Genético de Plantas da Universidade Federal Rural de Pernambuco, como requisito parcial à obtenção do título de Doutora em Agronomia, na área de concentração Melhoramento Genético de Plantas.

COMITÊ DE ORIENTAÇÃO:

Professor Dr. Reginaldo de Carvalho - Orientador - UFRPE

Professor Dr. Cícero Carlos de Souza Almeida - Coorientador - UFAL

RECIFE – PE

2022

**CARACTERIZAÇÃO GENÔMICA DO DNA REPETITIVO E DIVERSIDADE
GENÉTICA EM *Syagrus coronata* (Mart.) Becc.**

JACKELINE TERTO DA SILVA SANTANA

Tese defendida e aprovada pela Banca Examinadora em 30/06/2022

COMISSÃO EXAMINADORA:

Dr. Reginaldo de Carvalho (orientador)

Universidade Federal Rural de Pernambuco – UFRPE

Dra. Gheysa Coelho Silva (examinadora interna)

Universidade Federal Rural de Pernambuco – UFRPE

Dra. Iêda Ferreira de Oliveira (examinadora externa)

Universidade Federal Rural de Pernambuco – UFRPE

Dr. Jailson Gitaí dos Santos Frazão (examinador externo)

Universidade Federal Rural de Pernambuco – UFRPE/UAST

Dr. João Pacífico Bezerra Neto (examinador externo)

Universidade de Pernambuco – UPE

RECIFE - PE

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- S232c Santana, Jackeline Terto da Silva
Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc. /
Jackeline Terto da Silva Santana. - 2022.
110 f. : il.
- Orientador: Reginaldo de Carvalho.
Coorientador: Cicero Carlos de Souza Almeida.
Inclui referências e anexo(s).
- Tese (Doutorado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em
Agronomia - Melhoramento Genético de Plantas, Recife, 2022.
1. palmeira nativa. 2. ouricuri. 3. fração genômica repetitiva. 4. marcadores moleculares. 5. recursos
genéticos. I. Carvalho, Reginaldo de, orient. II. Almeida, Cicero Carlos de Souza, coorient. III. Título

CDD 581.15

(...) Be cheerful. Strive to be happy.

Max Ehrmann

Ao meu avô, **ALCINO VICENTE DA SILVA**

(em memória desde 30 de maio de 2022)

DEDICO

AGRADECIMENTOS

A Deus, por ser infinitamente perfeito em tudo que faz. Mesmo que eu tente elaborar o melhor texto, jamais serei capaz de expressar tamanha gratidão pelas bênçãos a mim concebidas. Por isso, limito-me a dizer que “seja feita a Vossa vontade”.

Aos meus orientadores, Dr. Reginaldo de Carvalho e Dr. Cícero Carlos de Souza Almeida, pela oportunidade de realizar este trabalho, por todos os ensinamentos e amizade. Obrigada por tudo!

À Dra. Iêda Ferreira, ao Dr. Jailson Gitaí, à Dra. Gheysa Coelho, à Dra. Luiza Suely, ao Dr. João Pacífico, à Dra. Genira Pereira e ao Dr. Reginaldo de Carvalho pela disponibilidade de fazer parte da comissão examinadora e muito contribuírem para a melhoria deste trabalho. Obrigada!

À Dra. Luiza Suely Semen Martins, à Dra. Ana Christina Brasileiro Vidal, à Dra. Angélica Virginia Valois Montarroyos e ao meu orientador Dr. Reginaldo de Carvalho por fazerem parte de minha banca de qualificação do Doutorado e muito contribuírem para melhoria do manuscrito. Gratidão!

À Universidade Federal Rural de Pernambuco, em especial ao Programa de Pós-graduação em Agronomia – Melhoramento Genético de Plantas (PPGAMGP). Tenho muito orgulho de fazer parte dessa família!

À Universidade Federal de Alagoas – *Campus Arapiraca*, mais precisamente à equipe do Laboratório de Recursos Genéticos (LARGE) pelo acolhimento, ensinamentos e momentos de descontração ao longo da realização desta pesquisa. Muito obrigada, Gleica Maria, Eliane, Elvia Jéssica, Ewerton Cavalcante, Suzziane Moraes, Tomáz, Maria Aparecida (Cida), Grazielle (Grazi) e Richard.

À Coordenação do Curso de Pós-graduação em Agronomia – Melhoramento Genético de Plantas (PPGAMGP), pela dedicação e paciência direcionadas ao corpo discente.

A todos os professores do PPGAMGP, Dr. Reginaldo de Carvalho, Dr. José Wilson, Dr. Gerson Quirino, Dr. Dimas Menezes, Dra. Vivian Loges, Dr. Francisco Rodrigues, Dr. Edson Ferreira, Dra. Gheysa Coelho, Dra. Luiza Semen, Dr. Péricles Albuquerque, Dr. Antônio Mendonça, Dra. Rosimar Musser, Dr. Roberto Albuquerque, Dr. José Luiz e Dra. Roseane Cavalcanti.

Aos meus colegas do PPGAMGP, principalmente, Sérgio Rogério, Gersia Gonçalves, Islan Diego, Djayran Sobral, Cristina Pereira, Antônio Costa, Jackson da Silva, Jamile Erica, Kleyton Danilo, Flavia Gomes, Bianca Galucio, Ricardo de

Normandes, Eduardo Lopes, Damião Queiroz, João Albuquerque, Ana Maria, Robson Ramos, Fabian Santana, Fernando Rocha, Cinthia, Agnes.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pela concessão da bolsa de estudos e à FAPEAL (Fundação de Amparo à Pesquisa do Estado de Alagoas) por financiar este trabalho.

A meus amados pais, José Izaías e Cícera Terto, por todo exemplo de amor, carinho, garra e persistência. A educação e ensinamentos que recebo de vocês diariamente são o melhor presente que um filho pode ter. Amo vocês!

Ao meu colega de profissão, melhor amigo e noivo, Sérgio Rogério, pelo apoio, companheirismo e paciência. Obrigada por caminhar lado a lado comigo nesta jornada. Amo você!

À Joelma Terto, querida e amada irmã, por incentivar a realização de cada um dos meus objetivos, dando-me os mais sábios conselhos.

À Alice Terto, sobrinha e afilhada que tanto amo. Eis a responsável pelos momentos de distração e diversão genuína durante a realização deste trabalho.

A todos os meus amigos, demais familiares e àqueles que, de certa forma, torcem por mim, meu mais profundo agradecimento.

A todos que, direta ou indiretamente, contribuíram para a realização desta pesquisa.

LISTA DE ABREVIATURAS

cDNA – Complementary DNA

CPD – Caatinga Phytogeographic Domain

cpDNA – chloroplast DNA

CRAD – Centro de Referência para Recuperação de Áreas Degradadas

CsCl – Cloreto de Césio

DNA – Deoxyribonucleic acid

IN – Instrução Normativa

ITR – Interstitial Telomeric Repeats

ITS – Internal Transcribed Spacer

IUCN – International Union for Conservation of Nature

Kb – Kilobase

LINE – Long Interspersed Nuclear Elements

LTR – Long terminal repeats

Mb – Megabase

NGS – Next Generation Sequencing

PCR – Polymerase chain reaction

Pg – Picogramas

PLE – Penelope

rDNA – ribosomal DNA

RNA – Ribonucleic acid

SatDNA – satellite DNA

SINE – Short Interspersed Nuclear Elements

SSR – Simple Sequence Repeat

TEs – Transposable elements

LISTA DE FIGURAS

Capítulo I

Figura 1. Mapa de distribuição geográfica de *Syagrus coronata* (Mart.) Becc. **Fonte:** NOBLICK (2017)..... 18

Figura 2. *Syagrus coronata* (Mart.) Becc. (ouricuri) **A.** População nativa. **B.** Inflorescência e espata coriárea. **C.** Infrutescência e **D.** Matriz com destaque para frutos maduros, disposição foliar e bainha de folhas velhas na base do estipe. **Fonte:** Terto, J..... 20

Figura 3. **A.** Frutos maduros e sementes secas de *S. coronata*. **B.** Sementes evidenciando o tegumento e amêndoas e **C.** Corte transversal e diagonal em sementes de ouricuri. **Fonte:** Terto, J.....22

Capítulo II

Figure 1. Flow cytometric histogram of relative fluorescence intensity of *Syagrus coronata* (A) and *Zea mays* CE-777 (B).....69

Figure 2. Repetitive DNA fraction in the genome of 101 plants, distributed in 37 Orders.....70

Figure 3. Different factors as function of repetitive frequency. (A) Genome size; (B) Retrotransposons; (C) Transposons and (D) Tandem Repeats.....71

Capítulo III

Figura 1. (A) Distribution of haplotypes from the *trnH-psbA*. The indicated colors indicate the different haplotypes found and e (B), (C) e (D) Allelic distribution in three microsatellite regions. The abbreviations specified refer to the localities: Loc. A (São José da Tapera), Loc. B (Pirambú), Loc. C (Coruripe), Loc. D (Santana do Ipanema), Loc. E (Raso da Catarina) and Loc. F (Água Branca).....92

Figura 2. Principal component analysis (PCA) revealing a two-dimensional distribution using data of SNP and microsatellites. Minimum-spanning network depicting the distribution multilocus SNP and microsatellites in *Syagrus coronata*.....93

Figura S1. Percent missing during polymerase chain reaction (PCR) in diferent loci.....94

Figura S2. Clustering obtained from microsatellite sequences (SSRs) using the Neighbor Joining method.....95

LISTA DE TABELAS

Capítulo I

Tabela 1. Classificação dos elementos transponíveis (TEs) com base em Wicker et al. (2007).....	39
--	----

Capítulo II

Table 1. Summary of repetitive DNA sequences in <i>Syagrus coronata</i>	72
--	----

Table 2. Characterization of satellite sequences in <i>Syagrus coronata</i>	73
--	----

Table S1. Characterization of monomers of satellite sequences in <i>Syagrus coronata</i>	74
---	----

Table S2. Species and families used to correlate different factors as a function of repetitive frequency.....	75
--	----

Capítulo III

Table 1. Genetic diversity parameters for <i>Syagrus coronata</i> using sequences data.....	96
--	----

Table 2. Analysis of Molecular Variance between samples using sequences data from chloroplast.....	97
---	----

Table 3. Genetic diversity parameters for <i>Syagrus coronata</i>	98
--	----

Table 4. Genetic diversity parameters for <i>Syagrus coronata</i>	99
--	----

Table 5. Analysis of Molecular Variance between samples using SSR data.....	100
--	-----

Table 6. Pair-by-pair matrix based on the values of genetic differentiation (F_{ST}) for the population groups of <i>Syagrus coronata</i> , where above contains the values F_{ST} from SSR data and below the F_{ST} from SNP data.....	101
---	-----

Table S1. Details of six populations of <i>Syagrus coronata</i> collected in the Caatinga and Atlantic Forest, indicating latitude and longitude.....	102
--	-----

SUMÁRIO

RESUMO GERAL	13
GENERAL ABSTRACT	14
CAPÍTULO I	14
1. INTRODUÇÃO GERAL.....	15
2. REFERENCIAL TEÓRICO	17
2.1. ASPECTOS GERAIS DA PALMEIRA OURICURI (<i>Syagrus coronata</i> (Mart.) Becc.)	17
2.2. DESCRIÇÃO MORFOLÓGICA	19
2.3. ATRIBUTOS ECOLÓGICOS E POTENCIAL SOCIOECONÔMICO	22
2.4. MELHORAMENTO GENÉTICO DA PALMEIRA OURICURI	26
2.5. CONSERVAÇÃO E RECURSOS GENÉTICOS DE OURICURI.....	27
2.6. DIVERSIDADE GENÉTICA E MÉTODOS ESTATÍSTICOS DE MEDIDA	30
2.7. MARCADORES MICROSSATÉLITES E SEQUÊNCIAS DE CLOROPLASTO ..	32
2.8. TAMANHO DO GENOMA VEGETAL	33
2.9. DNA REPETITIVO.....	36
2.9.1. SEQUÊNCIAS REPETITIVAS EM “TANDEM”	38
<i>SEQUÊNCIAS DE DNA SATÉLITE</i>	38
<i>DNA RIBOSSOMAL (DNAr)</i>	39
<i>DNA TELOMÉRICO</i>	40
2.10. SEQUÊNCIAS REPETITIVAS DISPERSAS	40
<i>RETROTRANSPOSONS (CLASSE I)</i>	41
<i>TRANSPOSONS (CLASSE II)</i>	42
3. REFERÊNCIAS BIBLIOGRÁFICAS	44
CAPÍTULO II	56
Genome size and <i>de novo</i> identification repetitive fraction in the <i>Syagrus coronata</i> (Arecaceae).....	56
CAPÍTULO III	85
Diversidade genética em <i>Syagrus coronata</i> (Mart.) Becc. (Arecaceae): implicações para manejo e conservação genética.....	85

RESUMO GERAL

Syagrus coronata (Mart.) Becc., popularmente conhecida como ouricuri e licuri, é integrante de Arecaceae e apresenta grande importância socioeconômica e ecológica. Apesar de ser uma espécie amplamente distribuída na Caatinga e Mata Atlântica, a expansão urbana e intensificação de práticas agropecuárias têm contribuído para o declínio populacional acentuado, necessitando de medidas que possam auxiliar a conservação de seus recursos genéticos. Estudos genômicos envolvendo a caracterização da fração repetitiva e determinação do tamanho do genoma vêm sendo realizados para espécies de Arecaceae. No entanto, para *S. coronata* há a necessidade de estudar a organização do DNA repetitivo. Diante disso, objetivou-se neste estudo: 1) estimar o tamanho do genoma e caracterizar o DNA repetitivo em *S. coronata* por citometria de fluxo e abordagem *k-mer* usando o método RESPECT e 2) analisar a diversidade genética de populações de *S. coronata* em dois biomas por meio de marcadores microssatélites e DNA de cloroplasto. O tamanho do genoma foi estimado em $2,27 \pm 0,026$ Gbp/1C usando citometria de fluxo e 2,76 Gbp usando abordagem *k-mer*. A fração repetitiva do genoma correspondeu a 60,39%, onde as sequências retrotransposons LTR foram mais abundantes (37,62% para Ty1/Copia e 12,2% para Ty3/Gypsy). Outros elementos repetitivos foram identificados: transposons (3,91%), DNA ribossômico (0,56%) e sequências satélites (1,32%). Houve elevada diversidade em número e comprimento de monômeros nas sequências satélites. Os resultados de diversidade genética em seis diferentes localidades dos biomas Caatinga e Mata Atlântica revelaram a formação de apenas um grupo populacional, indicando existir pouca diferenciação genética entre as localidades estudadas. A maior parte da variação genética esteve presente dentro das amostras (98,3% para dados de SSRs e 56,68% para SNPs), o que indica a presença de fluxo gênico dentro da população. Para fins de conservação da espécie, sugere-se principalmente a conservação *in situ* visto que a alta diversidade encontrada dentro das populações indica ampla distribuição nas diferentes localidades.

Palavras-chave: Palmeira nativa, ouricuri, fração genômica repetitiva, marcadores moleculares, recursos genéticos.

GENERAL ABSTRACT

Syagrus coronata (Mart.) Becc., popularly known as ouricuri and licuri, is a member of Arecaceae and has great socioeconomic and ecological importance. Despite being a species widely distributed in the Caatinga and Atlantic Forest, the urban expansion and intensification of agricultural practices have contributed to the accentuated population decline, requiring measures that can help the conservation of its genetic resources. Genomic studies involving the characterization of the repetitive fraction and determination of the genome size have been carried out for species of Arecaceae. However, for *S. coronata* there is a need to study the organization of repetitive DNA. Therefore, the objective of this study was: 1) to estimate the genome size and characterize the repetitive DNA in *S. coronata* by flow cytometry and *k-mer* approach using the RESPECT method and 2) to analyze the genetic diversity of populations of *S. coronata* in two biomes using microsatellite markers and chloroplast DNA. The genome size was estimated at 2.27 ± 0.026 Gbp/1C using flow cytometry and 2.76 Gbp using the *k-mer* approach. The repetitive fraction of the genome corresponded to 60.39%, where retrotransposons LTR sequences were more abundant (37.62% for Ty1/Copia and 12.2% for Ty3/Gypsy). Other repetitive elements were identified: transposons (3.91%), ribosomal DNA (0.56%) and satellite sequences (1.32%). There was high diversity in the number and length of monomers in the satellite sequences. The results of genetic diversity in six different localities of the Caatinga and Atlantic Forest biomes revealed the formation of only one population group, indicating that there is little genetic differentiation between the studied localities. Most of the genetic variation was present within the samples (98.3% for SSR data and 56.68% for SNPs), which indicates the presence of gene flow within the population. For species conservation purposes, *in situ* conservation is mainly suggested since the high diversity found within populations indicates wide distribution in different localities.

Keywords: Palm tree, ouricuri, repetitive genomic fraction, molecular markers, genetic resources.

CAPÍTULO I

Referencial teórico

1. INTRODUÇÃO GERAL

Syagrus coronata (Mart.) Becc. é uma espécie pertencente a Arecaceae, popularmente conhecida como “ouricuri” e “licuri” que possui grande relevância socioeconômica por apresentar ampla utilização para confecção de produtos para fins artesanais, alimentícios e para outros setores industriais (Noblick 2017). Além disso, trata-se de uma palmeira essencial para a fauna nativa que se alimenta de seus frutos, especialmente a ave *Anodorhynchus leari* (conhecida como ararinha azul), que enfrenta forte ameaça de extinção (Santos-Neto e Camandaroba 2008). A distribuição geográfica de ouricuri é considerada ampla, onde pode ser encontrada na Caatinga e também em domínios da Mata Atlântica (Noblick 2017, Souza et al. 2018).

As populações de *S. coronata* vêm enfrentando acentuado declínio devido, principalmente, à expansão urbana e substituição das áreas de ocorrência natural para a realização de práticas agropecuárias em desarmonia com os ecossistemas, resultando na elevada vulnerabilidade da espécie. Dessa maneira, é fundamental o desenvolvimento de estratégias que proporcionem a conservação dos seus recursos genéticos. Nesse sentido, estudos sobre diversidade genética por meio de marcadores microssatélites (SSR-*Simple Sequence Repeats*) em associação com sequências de DNA cloroplastidiais (cpDNA) são considerados promissores em decorrência da vasta distribuição dos SSRs no genoma, sendo altamente polimórficos, multialélicos, codominantes e de alta reprodutibilidade (Moura et al. 2017) e devido à abundância de sequências de cloroplasto no genoma, flexibilidade e facilidade no gerenciamento de dados (Thomson 2014).

Diante da diversificação genômica presente nas plantas, pesquisas têm sido realizadas para caracterizar a fração de DNA repetitivo, bem como classificar o tamanho genômico, que pode ser muito variável entre as espécies (Leitch et al. 2005, Neumann et al. 2019). O DNA repetitivo compreende os fragmentos homólogos de DNA que são encontrados em várias cópias no genoma, podendo representar até 90% do genoma em plantas (Mehrotha e Goyal 2014).

De maneira geral, estudos indicam sua participação na geração de variabilidade genética, contribuição na manutenção da integridade cromossômica, regulação gênica, dentre outros (Mestrovic et al. 2015, Zhu et al. 2016, Li et al.

2017), podendo apresentar um padrão de distribuição “*em tandem*” (sequências satélites de DNA, DNA ribossômico e DNA telomérico) ou dispersa ao longo do genoma (transposons de DNA e os retrotransposons) (Wicker et al. 2007, Mehrotha e Goyal 2014). Na atualidade, por meio do Sequenciamento de Nova Geração (NSG) e recursos avançados de bioinformática, é possível caracterizar a fração repetitiva do genoma (Weiss-Schneewiss et al. 2015).

Estudos genômicos vêm sendo realizados em espécies da família Arecaceae (Ibrahim et al. 2021), porém, em *S. coronata*, apesar de haver pesquisas genéticas (Souza et al. 2018, Áquila et al. 2018, Barbosa e Almeida 2018), trabalhos sobre a caracterização genômica precisam ser efetuados, pois podem corroborar para uma melhor compreensão acerca da organização da fração repetitiva e implicação no processo evolutivo do ouricuri e espécies correlacionadas. Além disso, considerando que a conservação e utilização dos recursos genéticos perfazem a base para programas de melhoramento genético, o conhecimento proveniente de estudos de diversidade genética em populações remanescentes de ouricuri pode auxiliar na elaboração de medidas que auxiliem a conservação de seus recursos genéticos.

2. REFERENCIAL TEÓRICO

2.1. ASPECTOS GERAIS DA PALMEIRA OURICURI (*Syagrus coronata* (Mart.) Becc.)

A família Arecaceae, única representante da Ordem Arecales, pertence ao grupo das Monocotiledôneas sendo constituída por cerca de 2.600 espécies distribuídas em mais de 260 gêneros tropicais e subtropicais (Lorenzi et al. 2004). De acordo com Lorenzi (2010), no Brasil, essa família, de distribuição ampla e habitats diversificados, possui cerca de 200 espécies distribuídas entre 43 gêneros, aproximadamente. Sendo que na região Nordeste do país são registradas cerca de 70 espécies pertencentes a 16 gêneros. Dentre eles, encontra-se o gênero *Syagrus* cujo número de espécies relatadas em estudos remotos equivale a 36 espécies, além de relatos de híbridos naturais (Noblick 1986, Lorenzi et al. 2004, Dransfield et al. 2005).

Atualmente, porém, sabe-se que *Syagrus* é constituído por 65 espécies (apenas 8 não são encontradas no Brasil), 2 subespécies e 14 híbridos naturais, com distribuição ampla na região Neotropical (Noblick 2017). Ainda segundo o autor, o Brasil recebe destaque por possuir, dentre outras, uma palmeira singular que tem a capacidade de sobreviver em regiões de extrema seca, denominada *Syagrus coronata* (Martius) Beccari. Trata-se de uma espécie endêmica do Brasil, mais precisamente do Domínio Fitogeográfico da Caatinga (CPD) (Lorenzi 2010, Noblick 2017) e Domínios da Mata Atlântica (Souza et al. 2018), configurando uma espécie altamente relevante nos ambientes onde é encontrada.

Syagrus coronata (Mart.) Becc. é, portanto, uma das espécies que integram a família Arecaceae, subfamília Arecoideae e tribo Coccoceae (Dransfield et al. 2005, APG IV 2016). Foi inicialmente descrita por Martius, em 1826, recebendo como basônimo *Cocos coronata* Martius, que foi alterado para *Cocos quinquefaria* Barbosa Rodrigues, no ano de 1900 e, em 1910, para *Glaziova treubiana* Beccari. Alguns anos depois, passou a ser especificada como *Syagrus quinquefaria* (Barbosa Rodrigues) Beccari e posteriormente como *Syagrus treubiana* (Beccari) Beccari. Até que em 1916 recebeu o nome de *Syagrus conorata* (Martius) Beccari (Noblick 2017), perfazendo seu epíteto até os dias atuais. Já seu nome popular recebe variações a depender da região onde é encontrada, podendo ser conhecida como ouricuri, licuri, aricuri, nicuri, alicuri, baba de boi, coqueiro cabeçudo, coqueiro dicori, licurizeiro,

nicuri e urucuri, dentre outros nomes que geralmente as comunidades tradicionais, extrativistas e agricultores familiares designam (Drumond 2007, Noblick 2017).

Em seu trabalho proveniente de décadas de pesquisa, Noblick (2017) enfatiza a predominância desta palmeira na Caatinga sendo relatada há anos, evidenciando sua ampla distribuição que vai desde o Norte de Minas Gerais, abrangendo os Estados de Sergipe, Alagoas e se estendendo até o Sul de Pernambuco (Figura 1). No norte da Bahia, as populações de ouricuri distribuem-se ao longo da costa, todavia, à medida que a vegetação começa a mudar a leste – isto é, deixa de ser uma região de transição com vegetação seca e passa a ser representada pela vegetação típica de floresta úmida – nota-se uma redução populacional. Ainda segundo o autor, *S. coronata* demonstra ainda melhor adaptação em áreas de transição entre a Caatinga e campos rupestres, podendo ser considerada, assim, como um componente importante de algumas das florestas estacionais semidecíduais a leste de São Francisco. Já ao norte de Salvador, as populações de ouricuri se estendem até a zona de vegetação da restinga (litoral) (Noblick 2017). No entanto, populações de *S. coronata* também podem ser encontradas em domínios da Mata Atlântica (Souza et al. 2018), ressaltando a distribuição da espécie em diferentes biomas.

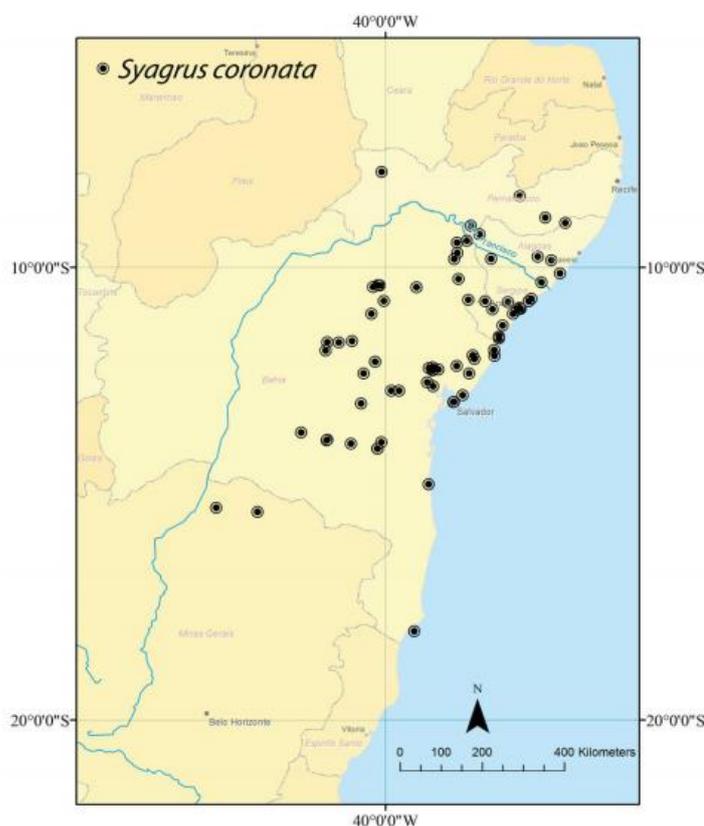


Figura 1. Mapa de distribuição geográfica de *Syagrus coronata* (Mart.) Becc. **Fonte:** NOBLICK (2017).

A palmeira ouricuri é caracterizada por ter ampla adaptação a diferentes condições do solo, apresentando bom desenvolvimento em solos férteis e profundos como também em solos pedregosos, podendo ainda ser encontrado em áreas com afloramentos rochosos. Entretanto, não possui adaptação em solos com elevada umidade (Drumond 2007).

2.2. DESCRIÇÃO MORFOLÓGICA

S. coronata faz parte da história do Brasil de maneira muito significativa, onde já era amplamente utilizada por povos indígenas. Relatos históricos provenientes dos séculos XVI e XVII enfatizam aspectos morfológicos peculiares a esta palmeira, como descrito a seguir:

*“As principais palmeiras bravas da Bahia são as que se chama de **ururucuri**, que não são muito altas e dão uns cachos de cocos muito miúdos, do tamanho e cor dos abricoques, aos quais se come o de fora, como os abricoques (...).”*

In: “Tratado descritivo do Brasil”, em que se diz das palmeiras ururucuri, de 1589, por Gabriel Soares de Souza, cuja reedição remonta ao ano de 1851, Recife: Fundação Joaquim Nabuco.

*“Dentre as palmeiras silvestres, as chamadas **Urycury** pelos indígenas, bem como o Tucu, não crescem em tanta altura (em comparação com as chamadas Pindova), são muito baixas que as outras e menos conhecidas. Eriçadas de espinhos aspérrimos; lenho férreo (...). Os frutos são menores que os da Pindova; no mais, quase da mesma natureza e sempre crescem em cachos.”*

In: “História Natural do Brasil Ilustrado”, capítulo X do Livro IV, em que se diz de palmeiras Urycury, de 1648, por Guilherme Piso, cuja reedição remonta ao ano de 1948, São Paulo: Companhia Editora Nacional.

Como relatado, esta palmeira possui uma estatura mediana, podendo atingir até 10 metros de altura, apresentando estipe único e ereto (a condição de perfilhamento não é muito comum) com aproximadamente 25 cm de diâmetro, cuja coloração é cinza escuro e com entrenós quase ausentes. Uma característica que merece destaque é que a porção superior do estipe geralmente encontra-se recoberta pela base persistente das bainhas das folhas mais velhas, conferindo-lhe um aspecto peculiar. Seu sistema radicular é do tipo fasciculado, sendo de

profundidade elevada e suas folhas são pinadas e de coloração verde (que varia de claro a intenso), podendo atingir até 3 metros de comprimento, possuindo projeções fibrosas similares a espinhos nas margens do pecíolo. As folhas são arranjadas em cinco fileiras encontradas comumente numa sequência em espiral – para a esquerda ou para direita – ao longo do estipe. Esta filotaxia na “coroa” refere-se justamente a seu epíteto atual “*coronata*” por formar esta copa frondosa com distribuição em fileiras (Figura 2 A e D) (Lorenzi et al. 2004, Drumond 2007, Noblick 2017).



Figura 2. *Syagrus coronata* (Mart.) Becc. (ouricuri) **A.** População nativa. **B.** Inflorescência e espata coriárea. **C.** Infrutescência e **D.** Matriz com destaque para frutos maduros, disposição foliar e bainha de folhas velhas na base do estipe. **Fonte:** Terto, J (2020).

Trata-se de uma espécie monoica, com inflorescência do tipo panícula medindo cerca de 35 – 55 cm de comprimento e de coloração amarela. Sua disposição é caracterizada por ser interfoliolar, ramificada em espiral e pendular sempre protegida por uma espata coriácea (Figura 2 B). Suas flores possuem um odor adocicado e são unissexuadas, configurando uma distribuição padrão, onde as flores pistiladas estão sempre entre duas estaminadas, o que denomina uma tríade (Lorenzi et al. 2004, Noblick 2017). A época de florescimento varia de acordo com a região, todavia pode-se afirmar que surgem predominantemente entre maio e agosto (Drumond 2007).

Os frutos estão arrançados em cachos, sendo que cada um apresenta um número estimado de 1.300 frutos, onde a unidade possui 2 cm de comprimento e 1,4 cm de diâmetro. O ouricuri frutifica e floresce durante o ano todo, porém, o pico de frutificação equivale ao período de dezembro a março. Em relação à coloração, apresentam variação de acordo com o estado de maturação, de modo que quando imaturos possuem coloração verde com o ápice amarronzado, no entanto, no processo de maturação adquirem uma nuance que permeia entre o amarelo ao laranja (Figura 2 C e D) (Crepaldi et al. 2001, Drumond 2007, Santos-Moura et al. 2016, Noblick 2017).

Ainda segundo os autores, os frutos são do tipo drupa com formato oval-elipsoide, apresentando exocarpo fibroso-tênuo, mesocarpo fibroso-mucilaginoso e endocarpo rígido. No processo de amadurecimento, o endosperma anteriormente líquido se solidifica. As sementes, quando secas, apresentam-se marrons e seu tegumento que é caracterizado pela elevada rigidez, revestindo a amêndoa. Esta é muito rica em óleo, tem coloração amarronzada na região externa, porém seu interior é branco (Figura 3 A, B e C). Em relação aos aspectos nutricionais, a amêndoa é muito rica em ferro, cálcio, cobre, magnésio, zinco, manganês, sais minerais e betacaroteno (Crepaldi et al. 2001).

A palmeira ouricuri possui um elevado potencial socioambiental e econômico nas áreas de ocorrência natural, sobretudo pelo fato de que todas as suas partes têm ampla utilidade (Campos et al. 2018). Andrade et al. (2015) registraram em suas pesquisas 537 usos conhecidos para *S. coronata*, onde aproximadamente 150 são praticados com mais frequência. Esta versatilidade atribuída à espécie representa uma fonte de renda não só para as populações humanas da região, mas também na

atualidade é vista como uma espécie notória para fins de mercado. Seu potencial está se destacando cada vez mais nos setores energético, alimentício, ornamental e paisagístico, fitoterápico, artesanal e forrageiro, além de configurar um papel de extrema importância ecológica.

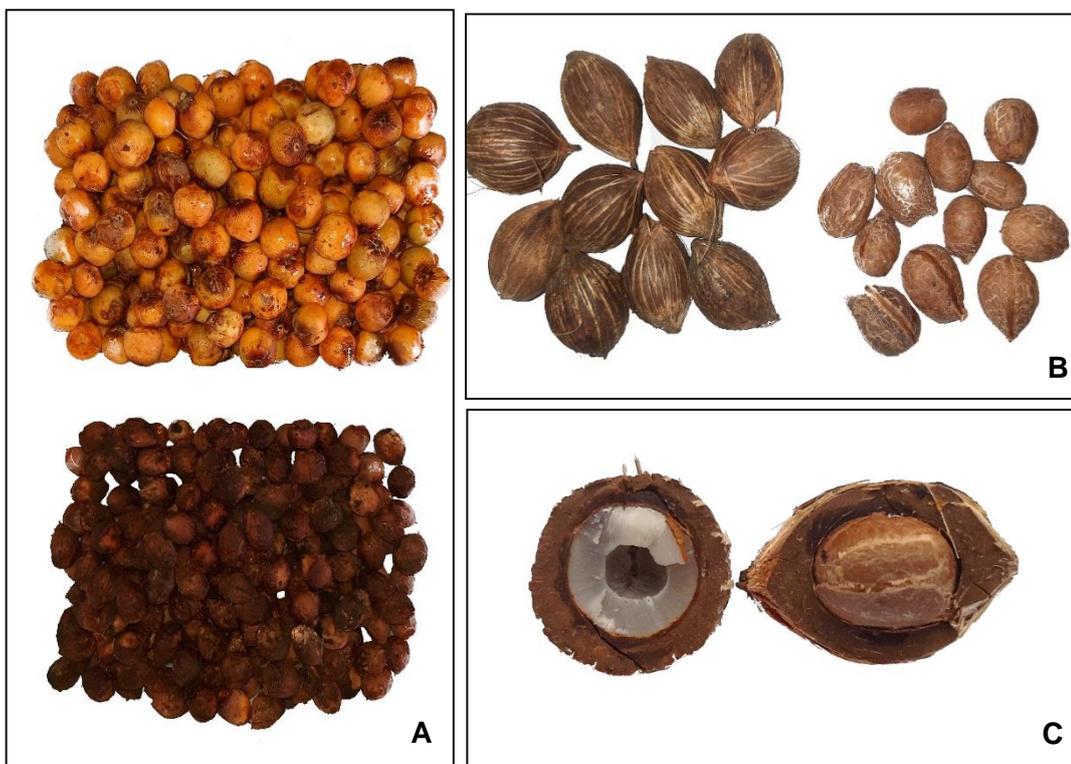


Figura 3. A. Frutos maduros e sementes secas de *S. coronata*. B. Sementes evidenciando o tegumento e amêndoas e C. Corte transversal e diagonal em sementes de ouricuri. **Fonte:** Terto, J (2020).

2.3. ATRIBUTOS ECOLÓGICOS E POTENCIAL SOCIOECONÔMICO

Do ponto de vista ecológico, configura uma espécie-chave nos ambientes onde é encontrada, sobretudo por constituir um dos alimentos mais importantes para a fauna nativa devido ao seu potencial de frutificação ao longo do ano (Drumond 2007). Dentre os animais que têm ouricuri como base alimentar, a ave *Anodorhynchus leari*, popularmente conhecida como arara-azul-de-Lear, merece destaque, pois seu principal alimento é justamente o fruto desta palmeira. Trata-se de uma ave endêmica do semiárido do Nordeste brasileiro, que se encontra ameaçada de extinção devido, sobretudo, ao tráfico e também pela redução cada vez mais elevada de seu alimento específico (IBAMA 2006, Santos-Neto e Camandaroba 2008, Lima et al. 2014).

S. coronata assume um papel importantíssimo no que se refere à associação com outras espécies vegetais popularmente conhecidas como “plantas caroneiras” ou epífitas que podem pertencer tanto ao grupo das briófitas como ao grupo das pteridófitas e angiospermas. Dentre todas elas, as espécies pertencentes à família Bromeliaceae, Cactaceae, Orchidaceae e Euphorbiaceae são comumente encontradas. De maneira geral, estas plantas fixam-se às bainhas foliares e também no estipe para sobreviverem, através da absorção de matéria orgânica e água. Esta relação simbiótica promove uma maior diversidade de animais dispersores de sementes (zooecia), contribuindo diretamente para a regeneração de áreas degradadas, evidenciando a importância da palmeira ouricuri como hospedeira da riqueza e diversidade destas epífitas (Oliveira et al. 2015, Carvalho et al. 2016, Castro et al. 2016).

Por ser uma planta de aspecto robusto e peculiar, ouricuri vem ganhando destaque como palmeira ornamental (Noblick e Barreto 2018). De acordo com Lorenzi et al. (2004) e Drumond (2007), a arquitetura de suas folhas atrelado à robustez de seu estipe, dentre outros fatores, chamam a atenção de profissionais da área do paisagismo de modo que a tendência de popularidade e utilização do ouricuri no mercado ornamental configura ascensão. Nesse sentido, o melhoramento genético de plantas torna-se fundamental para promover estudos de seleção de genótipos com características favoráveis a este uso, pois mesmo com boa aceitação pelo consumidor, pesquisas no âmbito paisagístico ainda são muito incipientes para esta palmeira.

O potencial alimentício do ouricuri, assim como suas características gerais, é comumente reportado na história do Brasil, onde já era descrito a sua utilização por povos indígenas como uma fonte de alimentação e subsistência, como abaixo mencionado:

“(...) e quebrando-lhe o caroço, donde se lhe tira um miolo como o das avelãs, que é alvo e tenro e muito saboroso, os quais coquinhos são mui estimados de todos. Estas palmeiras têm o tronco fofo, cheio de um miolo alvo e solto como o cuscuz, e mole; e quem anda pelo sertão tira esse miolo mole e cozi-o em um alguidar ou tacho, sobre o fogo, onde se lhe gasta a umidade, e é um mantimento muito sadio, substancial e muito proveitoso aos que andam pelo sertão (...).”

In: “Tratado descritivo do Brasil”, em que se diz das palmeiras urucuri, de 1589, por Gabriel Soares de Souza, cuja reedição remonta ao ano de 1851, Recife: Fundação Joaquim Nabuco.

Na atualidade, pode-se afirmar que muitas comunidades locais têm esta palmeira como uma fonte de alimentação e subsistência o que a torna um recurso essencial para estas famílias. O consumo do palmito oriundo do estipe é por vezes mencionado como uma forma de alimentação em substituição da farinha, por possuir um sabor adocicado e bem característico após cozido, triturado e peneirado. É verídico que este foi - e ainda é – utilizado por famílias de forma a combater a fome (Noblick 2017).

Entretanto, muito engana-se quem pensa em ouricuri apenas como uma fonte de alimentação secundária. De maneira geral, se bem manejado, um ouricurizal pode produzir até 4.000 kg/ha/ano, mesmo em condições adversas (Drumond 2007). Os frutos de ouricuri, mais especificamente suas amêndoas, são amplamente consumidas, tanto “*in natura*”, como na fabricação de sorvetes, doces, geleias, licores, compotas, cocadas e outros subprodutos (Carvalho et al. 2016), de modo que vêm ganhando cada vez mais notoriedade. A comercialização das amêndoas “*in natura*”, na maioria das vezes, segue um padrão de organização em forma de “rosário” (Noblick 2017), ou seja, as amêndoas são transpassadas por um cordão em fileiras assimilando-se a um colar e posteriormente comercializadas em feiras locais. Estudos acerca da composição nutricional dos frutos de ouricuri revelam que estes são altamente calóricos e suas amêndoas são constituídas principalmente por lipídeos e proteínas (Crepaldi et al. 2001), chegando a fornecer até 39% de óleo em sua constituição (Salles et al. 2010).

Por possuir grande concentração de lipídeos, a técnica de extração de óleo proveniente das amêndoas é considerada uma prática ascendente devido aos diversos benefícios atrelados. O óleo extraído, portanto, pode ser destinado para uso na culinária que, aliás, é muito usado pela população do semiárido (Andrade et al. 2015) e/ou na indústria para a fabricação de saponáceos de excelente qualidade (sabão em pó, detergentes, sabão em barra, sabonetes finos) (Noblick 2017).

O óleo possui ainda outra utilidade muito relevante para a humanidade, mediante a busca crescente por fontes alternativas de combustíveis, que é justamente seu potencial na produção de biodiesel (Carvalho et al. 2016). De acordo com Salles et al. (2010), *S. coronata* é uma fonte de triglicerídeos adequada para a síntese de biodiesel com grande potencial de mercado. Esta característica inclusive

tem despertado o interesse de vários produtores rurais a fim de manejar seus povoamentos naturais ou até mesmo de estabelecer novos plantios com esta finalidade (Drumond 2007). Todavia, vale a pena destacar a necessidade do desenvolvimento de pesquisas científicas para gerar mais informações, tornando esta espécie mais conhecida a nível nacional e também oferecer suporte técnico para melhor orientar os produtores rurais no cultivo.

A convivência com a palmeira ouricuri permitiu que as comunidades rurais desenvolvessem técnicas de aperfeiçoamento e utilização de suas folhas na geração de renda, tornando-a um recurso de subsistência essencial de modo que essa aptidão é repassada de geração a geração. Neste sentido, as folhas possuem uma maleabilidade peculiar que confere resistência e facilidade na fabricação de diversos produtos e, por isso, são amplamente usadas como fontes de matéria-prima para confecção artesanal, a citar: vassouras, redes de pesca, chapéus, bolsas, cestas, esteiras, espanadores, bijuterias, peneiras, dentre outros (Andrade et al. 2015, Noblick 2017). Tendo em vista este potencial, associações e cooperativas estão sendo criadas para promover uma melhor organização entre os agricultores familiares e gerar uma maior visibilidade perante ao público (a exemplo da Associação das Artesãs do Pontal do Coruripe – AL e Cooperativa de Produção da Região do Piemonte da Diamantina – BA), permitindo que seus produtos sejam cada vez mais inseridos no mercado.

Além disso, as folhas após serem trituradas podem ser usadas como complemento na alimentação animal, principalmente em períodos de escassez hídrica. Servem também como componente na fabricação de casebres e construções rurais (Rufino et al. 2008, Andrade et al. 2015, Noblick 2017) e através da raspagem de suas folhas pode-se extrair uma cera de excelente qualidade (similar à cera da carnaubeira) (Crepaldi et al. 2010) que é muito propícia para a fabricação de papel carbono, graxa para sapatos, tinta para pintura de móveis e automóveis (Lorenzi 2010).

S. coronata possui um amplo potencial fitoterápico, entretanto, ainda é pouco explorado e divulgado perante a sociedade. Todavia, seus benefícios são conhecidos e utilizados por populações locais de modo que os relatos de eficácia têm despertado de maneira crescente o interesse para o desenvolvimento de pesquisas. Dentre os efeitos fitoterápicos, pode-se mencionar o uso da água

proveniente dos frutos, mais especificamente do endosperma, para tratamentos oculares (colírio), bem como agente de combate a micoses, auxiliando também para promover uma cicatrização mais rápida em ferimentos; o uso das raízes é comumente citado como um ingrediente corriqueiro em chás para auxiliar o desequilíbrio na produção de hormônios no período de menopausa e com potencial calmante, além de ser usado para combater dores da coluna vertebral (Rufino et al. 2008). Pesquisas investigando as propriedades antimicrobianas de ouricuri foram realizadas e revelaram que, de fato, trata-se de uma espécie potencial para esta característica, principalmente devido a compostos presentes em suas inflorescências (Hughes et al. 2013).

É inquestionável o potencial atribuído à *S. coronata*. Todavia, as questões de cultivo (com adequado manejo), aperfeiçoamento tecnológico e, principalmente, a exploração consciente pelas comunidades locais configuram um crescimento produtivo e maior representação dos produtos diretos e indiretos, levando em consideração a consolidação no mercado. Faz-se necessário, portanto, um melhor planejamento de estratégias de aproveitamento da palmeira como um todo por meio de bases racionais e sustentáveis.

2.4. MELHORAMENTO GENÉTICO DA PALMEIRA OURICURI

De maneira geral, a formação de bancos de germoplasma e desenvolvimento de programas de melhoramento são essenciais para o processo de domesticação das espécies. Nesse sentido, o pré-melhoramento envolvendo as etapas de caracterização e obtenção de informações acerca da diversidade é essencial (Manfio et al. 2011).

A forma mais comum de exploração de *S. coronata* na atualidade é mediante o extrativismo, não havendo ainda programas de melhoramento genético, apesar de esta palmeira possuir elevada importância do ponto de vista agrícola devido ao vasto potencial de utilização para atender as necessidades da indústria como, por exemplo, nos setores paisagístico (Noblick e Barreto 2018), alimentício (Andrade et al. 2015), saponáceo (Noblick 2017), combustível alternativo (Carvalho et al. 2016). Dessa maneira, fazem-se necessários mais estudos que forneçam mais informações acerca de sua biologia reprodutiva, caracterização morfológica, capacidade produtiva, variabilidade genética, dentre outros fatores que são

utilizados como base para auxiliar no desenvolvimento de programas de melhoramento genético.

Pesquisas e publicações acerca da caracterização da biologia floral de *S. coronata* são incipientes, porém, demonstram que se trata de uma espécie xenógama facultativa, em que possui um sistema reprodutivo misto, não havendo auto-incompatibilidade. A polinização ocorre, principalmente, através de insetos (com destaque para espécies de abelhas e besouros) e do vento. Já a dispersão dos frutos se dá por meio da zoocoria, onde pode ser realizada por animais silvestres como também por animais domésticos que forrageiam em áreas de ocorrência da palmeira (Rocha 2009).

Embora em número reduzido, alguns estudos vêm sendo realizados com objetivo de avaliar características morfológicas e produtivas que podem ser utilizadas em futuros programas de melhoramento genético da palmeira ouricuri, a citar: caracterização dos frutos, altura da planta, número de folhas por planta, comprimento da folha, largura da folha, projeção da copa, diâmetro do estipe, número de cachos, número de estruturas florais, peso de cacho, número de frutos por cacho, peso do engaço, comprimento do cacho e peso de 10 frutos frescos, em que já foi constatada notável variabilidade genética entre os genótipos avaliados com potencial para inserção em programas de melhoramento (Santos-Moura 2016, Neves 2021).

2.5. CONSERVAÇÃO E RECURSOS GENÉTICOS DE OURICURI

Ao decorrer dos anos as populações nativas de ouricuri vêm enfrentando notável depauperamento nas suas áreas de ocorrência natural, sobretudo pelo uso de práticas agropecuárias em desarmonia com os ecossistemas, tais como, queimadas, supressão indiscriminada de vegetação nativa e o sobrepastoreio do gado (Rocha 2009, Conceição et al. 2018).

S. coronata possui uma característica singular constituindo uma espécie totalmente aproveitável, gerando grandes benefícios para as comunidades que sobrevivem por meio da exploração extrativista (Noblick 2017). Todavia, a prática de extração de frutos, folhas e estipe de maneira não sustentável desde épocas coloniais contribuiu efetivamente no drástico declínio populacional. Dessa forma, em meados da década de 1940 alguns órgãos de caráter federal e municipal

começaram a estabelecer medidas visando uma exploração mais racional e voltada para a conservação da espécie, sendo a promulgação do decreto – Lei de 17 de setembro de 1941 uma das primeiras medidas legais específicas culminando na criação do Serviço de Defesa do Licurizeiro (Rocha 2009).

Apesar disso, os licurizais nativos continuaram a ser prejudicados de modo que no ano de 1996 a União Internacional de Conservação da Natureza (IUCN) inseriu *S. coronata* em sua lista de espécies vegetais ameaçadas, categorizando-a como Espécie Vulnerável e recomendou o desenvolvimento de alternativas para auxiliar na recuperação das populações afetadas (Johnson 1996). Em um cenário mais recente, com o intuito de promover e reger medidas favoráveis à conservação da palmeira ouricuri, foi estabelecida a Instrução Normativa (IN) 191/2008 que, dentre seus preceitos, proíbe o corte nas áreas de sua ocorrência natural, além de orientar sobre a prática de coleta de frutos e folhas para fins diversificados (Normas Brasil 2008).

De acordo com Gomes et al. (2018), os Estados de Pernambuco, Sergipe e Bahia têm a presença de *S. coronata* em suas Unidades de Conservação (UC), sendo a espécie localizada e monitorada principalmente em Parques Nacionais (Catimbau/PE, Chapada da Diamantina/BA e Serra de Itabaiana/SE), Floresta Nacional (Ibura/SE), Estação Ecológica (Canudos/BA), Áreas de Proteção Ambiental (Morro do Urubu/SE e Serra Branca/BA), Área de Relevante Interesse Ecológico (Cocorocó/BA), Parque Estadual (Canudos/BA) e em Parques Municipais (Pituaçu/BA e Zoobotânico/BA). Estudos diversificados são realizados nestas áreas mencionando a presença de ouricuri como espécie representativa da vegetação como, por exemplo, levantamento e caracterização florística (Fernandes e Queiroz 2015, Santana et al. 2017) e trabalhos que demonstram a importância de *S. coronata* na conservação da diversidade de espécies epífitas (Castro et al. 2016).

Tendo em vista o status atual de populações naturais desta espécie, mesmo mediante ações de conservação *in situ*, faz-se necessário o desenvolvimento de medidas que permitam assegurar um estado adequado de conservação do ouricuri. Neste sentido, medidas de conservação *ex situ* têm adquirido um papel relevante, sobretudo a despeito da inserção de sementes dessa palmeira no banco de germoplasma do Centro de Referência Para a Recuperação de Áreas Degradadas (CRAD-UNIVASF) (Gomes et al. 2018), com o intuito de preservar as características

genéticas da espécie para utilização em pesquisas futuras. De acordo com Nass et al. (2008), a formação adequada dos bancos de germoplasma é fundamental para a identificação de potenciais genitores a serem utilizados em programas de melhoramento vegetal, além de ser uma valiosa ferramenta para a conservação dos recursos genéticos.

Faz-se necessário a implementação de ações de conservação, tendo em vista seus valores imensuráveis para a manutenção da biodiversidade local como também para a comunidade rural (Rocha 2009). No caso mais específico desta palmeira, pode-se destacar que se trata de uma planta que assume um papel vital no ciclo de vida de diversos organismos no ambiente em que é encontrada, pois, mesmo sob déficit hídrico, permanece perenemente verde e frutificando, garantindo o sustento de famílias locais e alimento para uma série de animais (Noblick e Barreto 2018). Sendo assim, estudos agronômicos visando, principalmente, o aperfeiçoamento de técnicas de propagação, cultivo e tratos culturais, estratégias para a obtenção de uniformidade e redução de tempo no processo germinativo caracterizam algumas das ações a serem realizadas que podem subsidiar para uma exploração mais consciente, podendo contribuir no processo de conservação da espécie (Conceição et al. 2018).

É perceptível a necessidade de adoção de medidas que visem a conservação dos recursos genéticos por meio da intensificação de fiscalização do cumprimento das Leis de preservação, assim como incentivar a realização de projetos que visem a produção de mudas e capacitação dos produtores das comunidades que geram renda através uso desta palmeira.

Com esse intuito, publicações têm sido realizadas enfatizando a adoção de práticas de manejo para o extrativismo sustentável (Aroucha e Aroucha 2013, Brasil 2014), demonstrando sua importância como hospedeira de espécies epífitas (Oliveira et al. 2015, Castro et al. 2016), avaliando técnicas que contribuam para uma germinação eficiente (Carvalho et al. 2005, Rodrigues et al. 2006) e crescimento inicial das plantas (Carvalho et al. 2006), caracterizando a composição nutricional dos frutos (Crepaldi et al. 2001) e a morfologia de frutos, diásporos, sementes, plântulas e mudas (Santos-Moura et al. 2016). No entanto, vale a pena mencionar a importância do desenvolvimento de pesquisas científicas que auxiliem na manutenção das populações naturais de *S. coronata*, sobretudo estudos voltados

para o mapeamento da espécie, fenologia, biologia reprodutiva, aperfeiçoamento de técnicas de conservação e diversidade genética para subsidiar a conservação dos recursos genéticos (Conceição et al. 2018).

2.6. DIVERSIDADE GENÉTICA E MÉTODOS ESTATÍSTICOS DE MEDIDA

É sabido que a conservação e a exploração dos recursos genéticos de determinada espécie vegetal constituem a base para o desenvolvimento de um programa de melhoramento (Frankel et al. 1995) e que o conhecimento acerca da diversidade genética pode contribuir para auxiliar na conservação de germoplasma (Duran et al. 2009).

Dessa forma, entende-se diversidade genética a variação encontrada em sequências genômicas pertencentes a indivíduos de uma mesma espécie, sendo esta diversificação genética constituinte de um conjunto de marcadores específicos (Hartl e Clark 2010), que se origina mediante eventos de mutações que são acumuladas ao longo dos processos evolutivos (Arnold e Kunte 2017).

Os bancos de germoplasma caracterizam-se por possuir variabilidade genética contendo acessos de diferentes tipos, como acessos silvestres, cultivares obsoletas ou aquelas provenientes do melhoramento propriamente dito (Borém e Miranda 2009). É de extrema importância a caracterização da diversidade genética, sobretudo para que haja a identificação de genótipos distintos para uso no melhoramento e para fins de conservação. Além disso, espécies vegetais impostas a atividades extrativistas com manejo inadequado estão mais propensas a sofrerem redução populacional e, portanto, necessitam de estudos genéticos que caracterizem a diversidade para que estratégias de conservação dos recursos genéticos sejam elaboradas da melhor maneira possível.

De maneira geral, espécies de plantas que possuem sistema misto de reprodução e mecanismos eficientes para dispersão do pólen e sementes, tendem a apresentar uma maior variação genética dentro das populações e uma variação genética reduzida entre elas (Zanettini e Cavalli 2003).

A determinação da diversidade genética intra e interpopulacional podem ser estimadas por meio de marcadores moleculares. O Teorema de Hardy-Weinberg é caracterizado por ser o modelo clássico utilizado para estudos de diversidade em populações infinitas, cujo cruzamento seja aleatório, estando na ausência de

mutação, migração e seleção. Dessa forma, é possível a obtenção de inferências estatísticas para dados sobre diversidade genética. Dentre os métodos estatísticos utilizados para caracterizar a diversidade genética, as estatísticas F , a diversidade de Nei (Nei 1978) e as análises de variância das frequências alélicas são frequentemente usados, além dos índices H (Shannon 2001), G (Stoddart e Taylor 1988) e λ (Simpson 1949).

O valor de F_{ST} pode variar de 0 a 1, em que $F_{ST}=1$ indica subpopulações homocigotas em sua totalidade com a presença de alelos distintos entre elas. Por outro lado, quando o valor de $F_{ST}=0$ indica não haver diferenciação genética entre as subpopulações, pois as frequências alélicas são idênticas entre elas. Diante disso, interpretações de valores de F_{ST} foram sugeridas para facilitar a interpretação, de modo que F_{ST} entre 0 – 0,05 confere pequena diferenciação genética; F_{ST} entre 0,05 – 0,15 confere diferenciação genética moderada; F_{ST} entre 0,15 – 0,25 confere grande diferenciação genética e F_{ST} com valor superior a 0,25 confere diferenciação genética elevada (Wright 1978).

A diversidade genética baseada em Nei (1978) configura um método capaz de obter a medida do grau de diferenciação genética entre pares de populações. Tal método leva em consideração as identidades de dois genes escolhidos aleatoriamente dentre e entre populações independentemente do número de alelos. Além disso, é um método de grande utilidade para uma análise rápida de uma grande quantidade de informações de frequências gênicas.

A análise de variância molecular (AMOVA) estima os componentes de variâncias análogos às estatísticas F e reflete a correlação da diversidade de haplótipos em diferentes níveis de subdivisão hierárquica. Sua base genética consiste nas distâncias genéticas como desvios da média de um grupo e usa os quadrados dos desvios como variâncias, o que permite a partição da variação genética entre e dentro das populações analisadas (Excoffier et al. 2002).

A estatística H (também conhecida como índice de Shannon) não se baseia na heterozigotidade da população. A partir do cálculo de diversidade, pode-se determinar o quanto a diversidade encontrada para um local difere da diversidade hipotética estabelecida (Jamhour e Laroca 2004). Já a estatística G , conhecida como índice de Stoddart e Taylor (1988), é utilizada para refletir a integração da diversidade genética de cada loci em grupos multilocos. A estatística λ ,

conhecida como índice de Simpson (1949), é um índice de dominância e reflete a probabilidade de dois indivíduos serem próximos geneticamente.

2.7. MARCADORES MICROSSATÉLITES E SEQUÊNCIAS DE CLOROPLASTO

Em estudos voltados para análise de populações, os marcadores moleculares codominantes são um dos mais indicados, visto que geralmente apresentam altos níveis de polimorfismo intraespecífico. Nesse sentido, o marcador SSR (*Simple Sequence Repeat*), também conhecido por microssatélite, se destaca pelo elevado polimorfismo apresentado e por possuir distribuição aleatória no genoma. Os microssatélites são sequências curtas de DNA que estão organizados lado a lado (em “*tandem*”) (Hodel et al. 2016, Moura et al. 2017), cuja variação do número de vezes em que o motivo se repete é proveniente de mutações ao longo da replicação do DNA, conhecidas como *slippage*, onde basicamente há a formação de sequências diferentes comprimentos quando a enzima DNA-polimerase comete um “erro” durante o processo de adição de pares de bases (Ferreira e Grattapaglia 1998).

Para analisar o polimorfismo encontrado nos marcadores SSRs, a técnica de reação em cadeia da polimerase (PCR) permite que haja a síntese *in vitro* de milhões de cópias de uma região específica de DNA através da atuação da enzima designada de Taq-DNA-polimerase. Neste processo, a dupla fita de DNA é desnaturada, dando origem a cadeias simples para que os *primers* específicos flanqueadores da região do microssatélite se liguem a regiões complementares da sequência alvo para que haja a síntese da nova fita de DNA em que, geralmente, essas etapas ocorrem por aproximadamente 30 ciclos o que permite a visualização em forma de bandas em gel de agarose (Ferreira e Grattapaglia 1998, Nybom 2004, Semagn et al. 2006).

Um dos grandes avanços para análise de genotipagem foi o desenvolvimento de fluorocromos usados para marcação de *primers* em microssatélites que, dentre suas vantagens, cita-se a possibilidade de usar mais de um marcador na mesma reação, desde que não haja sobreposição de tamanhos. Basicamente, o procedimento consiste na marcação da fluorescência na extremidade 5' que, a depender do tipo de fluoróforo utilizado, é excitado por um comprimento de onda emitindo uma luminescência, possibilitando a identificação dos fragmentos

amplificados e posterior análise em sequenciador automático (Kimpton et al. 1993, Missiaggia e Grattapaglia 2006).

Outro tipo de marcador muito utilizado em estudos de diversidade genética baseia-se em regiões específicas do genoma de cloroplasto (cpDNA). De maneira geral, o genoma de cloroplasto apresenta estrutura circular sendo representado por duas regiões: a LSC (*Large Single Copy Region*) e a SSC (*Small Single Copy Region*) interrompidas por duas cópias de IRa e IRb (*Large Inverted Repeats*) (Shinozaki et al. 1986). A presença de uma mutação pontual em sequências de cloroplasto, substituindo um nucleotídeo por outro, é caracterizado como SNP (*Single Nucleotide Polymorphism*). O uso destes marcadores moleculares tem permitido o aumento do conhecimento sobre diversidade genética (Morgil et al. 2020) sobretudo por sua abundância no genoma, facilidade no gerenciamento de dados e flexibilidade (Thomson 2014).

Dentre as maneiras para a detecção de SNPs, está a técnica baseada no sequenciamento de regiões específicas de DNA provenientes de amplificação via PCR, utilizando *primers* correspondentes a estas sequências (Morgil et al. 2020). Considerando que os SNPs se encontram amplamente distribuídos ao longo do genoma, podendo estar presentes em regiões codificadoras e não codificadoras, é relevante a universalidade e qualidade dos *primers* (Hollingsworth et al. 2011). Em plantas, o espaçador intergênico *trnH-psbA* (com comprimento de 200–500 pb) apresenta nível satisfatório de amplificação e sequenciamento, além de possuir alta variação de nucleotídeos (Storchová e Olson 2007).

2.8. TAMANHO DO GENOMA VEGETAL

As primeiras definições sobre o termo “tamanho do genoma” indicavam referência ao conteúdo de DNA pertencente ao conjunto monoploide (Hinegardner 1976) enquanto que o “valor C” (C-value) era frequentemente empregado como o conteúdo de DNA de todo complemento cromossômico ou cariótipo. Devido à necessidade de melhor esclarecer as expressões como também promover um conceito unificado e sem ambiguidades, várias discussões foram sendo construídas para descrever o conteúdo de DNA nuclear com mais facilidade (Greilhuber et al. 2005). Atualmente, tamanho do genoma é definido no sentido mais amplo e refere-

se à quantidade de DNA contida no núcleo celular haploide, convencionalmente intitulado de valor C (Pellicer et al. 2018, Blommaert 2020).

Para estimar o conteúdo de DNA nuclear, inicialmente foram adotadas técnicas baseadas em análise química e também cinética de reassociação utilizadas para diversos tipos de organismos (Schmidt e Thannhauser 1945, Britten e Kohne 1968), porém, o fato de ambos os métodos serem realizados por meio de análises provenientes de um elevado número de células que, na maioria das vezes, estavam em fases de divisão celular divergentes e, portanto, com variações de quantidade de DNA, ocasionava estimativas de baixa precisão.

Nesse sentido, a técnica de microdensitometria de coloração por coloração Feulgen foi utilizada justamente por apresentar uma maior confiabilidade na estimativa na quantidade de DNA em relação às técnicas anteriores. Todavia, a necessidade de uma melhor comparação entre os resultados alcançados e também por conta de o tempo para obtenção dos dados ser elevado, além de ser uma técnica que demanda muito trabalho (Michaelson et al. 1991, Greilhuber et al. 2005), o uso desta técnica foi decaindo significativamente.

Atualmente, em plantas os métodos mais utilizados para estimar o tamanho do genoma são a abordagem baseada em *k-mers* (fragmentos de leituras com quantidade fixa de nucleotídeos) e a citometria de fluxo. O primeiro método consiste basicamente no número total de *k-mers* dividido pela cobertura do sequenciamento (Li e Harkess 2018) e o segundo permite a análise de dezenas de milhares de partículas de acordo com suas propriedades físicas e químicas em um curto período de tempo por meio do equipamento designado citômetro. Nesta técnica, as partículas em suspensão (células ou núcleos) são coradas com marcadores fluorescentes (fluorocromos) e atravessam um tubo capilar transparente que recebe um feixe de luz laser de comprimento de onda conhecido. Inicialmente, as partículas absorvem a luz e em seguida emitem uma faixa de luz com um comprimento de onda maior que é captada e transferida para um computador ligado ao sistema para ser interpretada (Bonnevier et al. 2018). Para obtenção de resultados eficientes é de extrema importância o preparo da suspensão de maneira adequada conforme descrito em protocolo (Loureiro et al. 2007). Dentre as principais vantagens deste método de determinar o tamanho do genoma em espécies vegetais, destaca-se a

alta reprodutibilidade e precisão (apresentando valores de coeficiente de variação menores que 5%), fácil preparo da amostra e rapidez (Souza et al. 2004).

A quantidade de DNA (valor C) é medida em picogramas (pg) (Blommaert 2020) e a depender do valor obtido, os genomas podem pertencer a categorias diferentes em função da variação de seus tamanhos. De acordo com Leitch et al. (2005), o tamanho dos genomas pode variar entre “muito pequenos” ($1C \leq 1,4$ pg), “pequenos” ($1C \leq 3,5$ pg), “intermediários” ($1C > 3,5$ a $< 14,0$ pg), “grandes” ($1C$ de $\geq 14,0$ a $< 35,0$ pg) e “gigantes” ($1C \geq 35,0$ pg).

Em plantas, a variação no tamanho genômico é bastante expressiva podendo ser facilmente observada em angiospermas, onde o menor genoma encontrado até então é de 0,07 pg pertencente a *Genlisea aurea* var. *aurea* A. St.-Hil. e *Genlisea tuberosa* Rivadavia, Gonella & A. Fleischm. (Lentibulariaceae) e o maior genoma medindo 152,23 pg pertencente a *Paris japonica* (Franch. & Sav.) Franch. (Melanthiaceae) (Pellicer et al. 2010, Pellicer et al. 2018, Leitch et al. 2019).

De acordo com o banco de dados do *Royal Botanic Garden Kew* (Londres), apesar do valor do tamanho de *P. japonica* ser considerado gigantesco em relação as demais angiospermas, dentre as 10.770 espécies pertencentes a esse grupo, a maioria apresenta genomas de tamanho $5,13 \pm 8,94$ pg (Leitch et al. 2019). É importante destacar que a variação de tamanho genômico não ocorre somente entre grandes níveis taxonômicos, podendo ocorrer também entre gêneros (Kelly et al. 2015, Vaio et al. 2016) e ainda entre populações dentro de uma mesma espécie, como é o caso do milho (*Zea mays* ssp. *mexicana*) (Bilinski et al. 2018).

Com base na informação de que existem genomas com tamanhos variados entre as espécies vegetais, um dos maiores impasses dentro da genômica é buscar cada vez mais compreender quais são os mecanismos responsáveis por esta variação e suas implicações. O entendimento da diversidade genômica apresentada entre os diferentes grupos de plantas terrestres pode auxiliar numa melhor interpretação acerca da evolução, além de impulsionar o aperfeiçoamento de técnicas moleculares. De maneira geral, sabe-se que os genomas podem apresentar uma expansão ou redução (Schubert e Vu 2016, Pellicer et al. 2018).

O evento denominado de poliploidia, ou seja, a duplicação do genoma completo, é por vezes uma das principais razões para promover o aumento do tamanho genômico e também uma das causas mais conhecidas, promovendo a

duplicação de genes e também o conteúdo de DNA não gênico (Bennetzen 2002). Todavia, alguns trabalhos têm demonstrado que há um processo denominado “genome downsizing” que ocorre após a poliploidia e pode acarretar na perda de DNA, diminuindo o tamanho do genoma (Leitch e Bennet 2004), mesmo que em outras pesquisas haja evidências de que em alguns casos ocorre o impedimento da perda de alguns genes (Pellicer et al. 2018). Outro evento que favorece a diferenciação no tamanho genômico é o ganho e a perda de cromossomos, evento definido como disloidia ascendente via fissão cromossômica e disloidia ascendente via fusão cromossômica, respectivamente (Schubert e Lysak 2011).

No entanto, o acúmulo de sequências repetitivas de DNA, principalmente transposons de DNA e retroelementos, configura um dos principais mecanismos contribuintes para o aumento do tamanho genômico (Mehrotra e Goyal 2014, Pellicer et al 2018), além de sequências a satélites (Garrido-Ramos 2015). De maneira geral, a diferenciação do tamanho do genoma por meio de DNA repetitivo tem sido alvo de muitas pesquisas com perspectivas distintas com o intuito de melhor compreender como estas sequências estão organizadas ao longo do genoma. É de fundamental importância que espécies pertencentes a grupos taxonômicos variados sejam estudadas, enfatizando aquelas cujos dados genômicos são considerados limitados (Pellicer et al. 2018).

Em Arecaceae, o número de pesquisas acerca da determinação do tamanho do genoma está em ascensão, porém o número de publicações ainda é considerado incipiente. No entanto, uma espécie que já tem dados publicados é a *Cocos nucifera* (popularmente conhecida como coqueiro), trata-se de uma espécie cujo gênero é próximo a *Syagrus* (Meerow et al. 2009), de modo que trabalhos demonstram seu genoma considerado pequeno, haja vista a classificação de Leitch et al. (2005) usando metodologias diferentes. Na literatura, os seguintes valores foram designados para o tamanho do genoma de *C. nucifera*: 2,42 Gpb (Xiao et al. 2017, Yang et al. 2021); 2,23 Gpg (Lantican et al. 2019) e 2C=5,48 pg (Freitas Neto et al. 2016).

2.9. DNA REPETITIVO

O termo “DNA repetitivo” compreende a fragmentos homólogos de DNA que são encontrados em várias cópias no genoma, sendo que a fração constituída por

estas sequências pode representar até 90% do tamanho do genoma em plantas (Mehrotra e Goyal 2014). O conhecimento sobre sequências repetitivas vem crescendo muito ao longo dos anos e cada vez mais a necessidade em estudá-las fica em evidência. Entretanto, nem sempre foi assim, em determinado momento a fração repetitiva do genoma foi considerada como “DNA lixo” e “elementos egoístas” (Orgel e Crick 1980, Doolittle e Sapienza, 1980).

Na atualidade, porém, sabe-se que o DNA repetitivo está relacionado à geração de variabilidade, contribuição para a manutenção da integridade dos cromossomos (podendo também auxiliar na orientação dos cromossomos durante o processo de divisão celular) e regulação da expressão de genes, além de algumas pesquisas evidenciarem sua participação no processo de especiação em plantas (Belyayev e Raskina 2010, Mestrovic et al. 2015, Biscotti et al. 2015, Zhu et al. 2016, Li et al. 2017).

Nesse sentido, diversos trabalhos têm sido realizados visando a caracterização do DNA repetitivo, localização, organização e abundância dessas sequências ao longo do genoma. O sequenciamento de Nova Geração (NGS) em associação com a bioinformática têm permitido melhor entendimento da natureza e comportamento dessa fração repetitiva (Weiss-Schneeweiss et al. 2015, Biscotti et al. 2015). Dentre as ferramentas utilizadas para identificação e classificação de sequências repetitivas, destaca-se a plataforma online e de livre acesso denominada *RepeatExplorer* que, por meio de similaridade, tem a capacidade de agrupar as sequências originando diferentes *clusters* que, por sua vez, representam diferentes famílias de DNA repetitivo (Novak et al. 2010, 2017, Garrido-Ramos 2017).

Há dois tipos de padrão de repetição destas sequências, podendo ser encontradas de maneira dispersa ao longo do genoma ou em “*tandem*” (dispostas lado a lado), sendo constituída por sequências satélites, DNA ribossomal (rDNA), DNA telomérico. Já a categoria dispersa possui a característica de possuir elementos genéticos com capacidade de realizar transposição, por isso são denominados de elementos transponíveis (TEs). Há dois tipos de TEs de acordo com seu modo de transposição, os transposons (transposição mediada por DNA) e retrotransposons (transposição mediada por RNA) (Kidwell 2002, Wicker et al., 2007, Mehrotra e Goyal 2014).

Em Arecaceae, mediante o número reduzido de publicações encontradas, a espécie *C. nucifera* apresentou 72,75% do seu genoma composto por DNA repetitivo, sendo que desse total, 92,23% correspondeu a sequências TEs (Yang et al. 2021). Resultado semelhante para esta espécie foi encontrado por Ibrahim et al. (2021), em que a fração repetitiva para o coqueiro foi 80,25%. Outra palmeira onde a fração de DNA repetitivo foi estabelecida foi a *Phoenix dactylifera* (conhecida como tamareira), sendo o valor correspondente a 41,42% (Ibrahim et al. 2021).

2.9.1. SEQUÊNCIAS REPETITIVAS EM “TANDEM”

SEQUÊNCIAS DE DNA SATÉLITE

As sequências repetitivas localizadas em blocos nos cromossomos, ou seja, que se encontram em uma localização específica com uma configuração em *tandem*, possuem diferentes comprimentos das unidades de repetição, bem como matrizes de tamanhos variados. Sendo assim, essas sequências podem ser classificadas em diferentes grupos a depender do seu tamanho e localização (Mehrotra e Goyal 2014). De acordo com Garrido-Ramos (2017), o termo “DNA satélite” engloba uma ampla diversidade de sequências representantes de uma parte da variação presente no genoma. As sequências de DNA satélite desempenham importantes papéis na regulação da cromatina, regulação de genes, auxiliando nos processos de segregação e preservação do material genético. Ainda segundo o autor, a diferenciação entre as famílias de SatDNA ocorre devido a localização, unidade de repetição, comprimento, abundância e sequência dos nucleotídeos.

A descoberta do DNAsat ocorreu através da ultracentrifugação, de modo que por meio de gradientes de densidade de cloreto de cério (CsCl), a fração de sequências satélites foi identificada em um gradiente como uma banda separada do resto do DNA (Kit 1961). Caracterizam-se por apresentar unidades de repetição ricas em AT, podendo formar matrizes de até 100 Mb. Os monômeros possuem comprimento que varia entre 150-400 pb (Mehrotra e Goyal 2014).

As matrizes de DNAsat são geralmente encontradas em heterocromatina (nas regiões pericentroméricas e subteloôméricas dos cromossomos), servindo como componente fundamental na estrutura cromossômica como, por exemplo, os

centrômeros e telômeros, além disso, as sequências satélites são consideradas a fração mais repetitiva do genoma (Garrido-Ramos 2017, Belyayev et al. 2019) e eventualmente podem contribuir de maneira efetiva para o aumento do tamanho do genoma (Garrido-Ramos 2015).

De maneira usual, as sequências repetidas em “*tandem*” denominadas minissatélites tem o comprimento do motivo de repetição variando entre 6-100 pb, enquanto que as matrizes podem medir entre 0,5-30 kb. Enquanto que os microssatélites possuem motivos de repetição bem menores (2-5 pb) (Mehrotra e Goyal 2014). Conforme revisado por Garrido-Ramos (2017), eventualmente pode ocorrer sobreposição entre essas sequências, por esta razão uma maneira de evitar esse problema é levando em consideração a localização, de modo que DNAsat configuram longos arranjos de heterocromatina, os minissatélites e os microssatélites perfazem regiões eucromáticas.

DNA RIBOSSOMAL (DNAr)

A síntese de proteínas é resultado da associação de uma diversidade de classes proteicas atrelado aos ribossomos (Neves et al. 2005). Para a maioria dos eucariotos, o DNA ribossomal é formado por dois tipos de genes que são denominados de DNAr 45S e 5S, sendo que o 45S é constituído pelas três regiões codificantes (18S, 5,8S e 26S), ou seja, a fragmentação desta molécula maior (45S) resultou na origem das demais.

O loco do DNAr é formado por unidades repetidas em *tandem* que são intercaladas por regiões de DNA espaçador chamados de ITS1 e ITS2 (*Internal Transcribed Spacer*) (Gruendler et al. 1991, Eickbush e Eickbush 2007). Embora para plantas às vezes o termo 45S seja usado, os genes para 18S, 5,8S e 25S são transcritos por um precursor de aproximadamente 35S (Volkov et al. 2004). Estudos iniciais em espécies vegetais já revelavam que milhares de cópias dessas sequências podem estar presentes ao longo do genoma de modo que podem representar um valor aproximado de 10% (Badaeva et al. 1996). O DNAr 5S, por sua vez, é formado por repetições de uma região codificante (aproximadamente 120 pb) que são intercaladas por DNA espaçador não transcrito (*Non Transcribed Spacer*). Esta região espaçadora caracteriza-se por ser bastante variável em número de sequências e tamanho (Douet e Tourmente 2007).

De acordo com Roa e Guerra (2012, 2015) em angiospermas, a maior frequência de DNAr 45S nos cromossomos foi encontrada na região terminal de braço curto, enquanto DNAr 5S mostraram-se mais frequentes na região proximal do braço curto. Devido as suas posições características nos cromossomos, estas sequências podem ser utilizadas como marcadores para identificação cromossômica (Moraes et al. 2007).

DNA TELOMÉRICO

Os estudos iniciais para estabelecer o motivo comum da sequência telomérica em plantas foram realizados em *Arabidopsis thaliana*, revelando que o motivo comum era constituído por TTAGGG (Richards e Ausubel 1988), que é a sequência encontrada para a maioria das espécies vegetais.

As sequências de DNA telomérico são encontradas nas extremidades dos cromossomos e atuam diretamente no processo de prevenção do encurtamento cromossômico durante a divisão celular (Kupiec 2014). Na literatura, porém, são descritos alguns mecanismos que pressupõem o mau funcionamento das enzimas DNA polimerase e telomerase durante a replicação do DNA (Lin e Yan 2008) e também indícios de fusão cromossômica no processo evolutivo de alguns grupos de plantas (Guerra 2008) que ocasionaram a identificação de sequências teloméricas intersticiais (ITR, *Interstitial Telomeric Repeats*) (Souza et al. 2016).

2.10. SEQUÊNCIAS REPETITIVAS DISPERSAS

Além das sequências repetitivas organizadas lado a lado, o genoma apresenta sequências de DNA dispersas, também conhecidos como elementos transponíveis (TEs). Estes componentes do genoma foram descritos primeiramente por Barbara McClintock, em que realizou estudos acerca da variação da cor dos grãos de milho (*Zea mays* L.) e notou a presença de “elementos” que tinham a capacidade de se mover ao longo do cromossomo para um local diferente, podendo afetar a expressão da pigmentação (McClintock 1950).

Estes elementos transponíveis equivalem a sequências de DNA e são um dos principais fatores responsáveis pela variação do tamanho genômico em espécies vegetais (Bennetzen et al. 2005, Lee e Kim 2014), podendo representar até 85% do genoma, como no caso do milho (Chénais et al. 2012). Os elementos transponíveis

são descritos como sequências importantes para a manutenção dos centrômeros e telômeros, podendo ter papel relevante no processo de regulação gênica, organização genômica (Kidwell 2002, Kazazian Jr 2004, Feschotte 2008, Rebollo et al. 2012). Devido à capacidade de inserção em lugares diferentes no genoma, os TEs apresentam uma notável variação e podem estar relacionados à ativação e/ou silenciamento de genes, além de poder favorecer a formação de novos rearranjos cromossômicos (Lisch 2013).

Desde a descoberta dos TEs, diversos pesquisadores se dedicaram para classificar esse DNA repetitivo disperso com base no mecanismo de transposição, composição enzimática e semelhança de sequências (Kidwell 2002, Wicker et al. 2007, Neumann et al. 2019). Dessa maneira, os TEs são separados em duas classes de acordo com seu modo de transposição: *Classe I* (composta por retrotransposons) e *Classe II* (constituída por transposons), com níveis hierárquicos menores (Wicker et al. 2007, Piégu et al. 2015).

É possível classificar as sequências em ordens, superfamílias e famílias. As ordens separam os elementos de acordo com a forma que estão organizados e mediante as particularidades enzimáticas e as superfamílias com base na similaridade de sequências. Já as famílias são organizadas a partir de sequências que possuem alta similaridade nas regiões codificantes ou nas repetições terminais (Wicker et al. 2007).

RETROTRANSPOSONS (CLASSE I)

Os retrotransposons, também denominados como retroelementos, se transpõem por meio de um mecanismo conhecido como “copia e cola” (*copy-and-past*) que promove o aumento do tamanho do genoma a cada evento de transposição, isto é, estas sequências se movem por meio de um RNA oriundo de sua sequência. Este RNA é traduzido para um DNA complementar (cDNA) através de uma enzima designada de transcriptase reversa e, posteriormente, o cDNA é transportado (por outra enzima chamada transposase), havendo o reconhecimento do sítio de inserção e fixação da cópia dos retroelementos (Wicker et al. 2007).

Os retrotransposons são classificados em cinco ordens: LTR (*Long terminal repeats*), DIRS, PLE (*Penelope-like*), LINE (*Long Interspersed Nuclear Elements*) e SINE (*Short Interspersed Nuclear Elements*) (as quatro últimas são conhecidas

como não LTRs). Há cinco superfamílias pertencentes à ordem LTR (*Copia*, *Gypsy*, *Bel-Pao*, *Retrovirus* e *ERV*), três superfamílias para DIRS (*DIRS*, *Ngaro*, *VIPER*), uma para PLE (*Penelope*), cinco para LINE (*R2*, *RTE*, *Jockey*, *L1* e *I*) e três para SINE (*tRNA*, *7SL* e *5S*) (Tabela 1) (Wicker et al. 2007).

TRANSPOSONS (CLASSE II)

As sequências repetitivas pertencentes à Classe II são conhecidas como transposons ou DNA transposons. Consistem em sequências que se transpõem por meio do mecanismo denominado “corta e cola”, em que a enzima transposase é a responsável por cortar o DNA e levar a sequência para outra posição dentro do genoma. Dessa forma, diferentemente do que ocorre com os retrotransposons, o número de cópias tende a ser menor. Todavia, em algumas situações pode ocorrer de determinados elementos conseguirem aumentar seu número de cópias. Por isso, os DNA transposons são subdivididos em duas subclasses. A subclasse I é caracterizada por transposons que não têm a capacidade de aumentar o número de cópias e subclasse II é constituída por sequências transposons que podem mover-se a partir de cópias geradas (Wicker et al. 2007).

A subclasse I possui duas ordens: TIR e Crypton, a primeira possui nove superfamílias (*Tc1-Mariner*, *hAT*, *Mutator*, *Merlin*, *Transib*, *P*, *PiggyBac*, *PIF-Harbinger* e *CACTA*) e a segunda possui apenas 1 (denominada *Crypton*). A subclasse II também possui duas ordens, que são Helitron (com apenas uma superfamília também denominada *Helitron*) e Maverick (com apenas uma superfamília também denominada *Maverick*) (Tabela 1) (Wicker et al. 2007).

Tabela 1. Classificação dos elementos transponíveis (TEs) com base em Wicker et al. (2007).

Classificação			
Classe	Ordem	Superfamília	Ocorrência
Classe I (Retrotransposons)	LTR	<i>Copia</i>	Plantas, Metazoários, Fungos e outros
		<i>Gypsy</i>	Plantas, Metazoários, Fungos e outros
		<i>Bel-Pao</i>	Metazoários
		<i>Retrovirus</i>	Metazoários
		<i>VER</i>	Metazoários
	DIRS	<i>DIRS</i>	Plantas, Metazoários, Fungos e outros
		<i>Ngaro</i>	Metazoários e Fungos
		<i>VIPER</i>	Outros
	PLE	<i>Penelope</i>	Plantas, Metazoários, Fungos e outros
	LINE	<i>R2</i>	Metazoários
<i>RTE</i>		Metazoários	
<i>Jockey</i>		Metazoários	
<i>L1</i>		Plantas, Metazoários, Fungos e outros	
SINE	<i>I</i>	Plantas, Metazoários e Fungos	
	<i>tRNA</i>	Plantas, Metazoários e Fungos	
	<i>7SL</i>	Plantas, Metazoários e Fungos	
	<i>5S</i>	Metazoários e outros	
Classe II (DNA transposons) Subclasse I	TIR	<i>Tc1-Mariner</i>	Plantas, Metazoários, Fungos e outros
		<i>hAT</i>	Plantas, Metazoários, Fungos e outros
		<i>Mutator</i>	Plantas, Metazoários, Fungos e outros
		<i>Merlin</i>	Metazoários e outros
		<i>Transib</i>	Metazoários e Fungos
		<i>P</i>	Plantas e Metazoários
		<i>PiggyBac</i>	Metazoários e outros
		<i>PIF-Harbinger</i>	Plantas, Metazoários, Fungos e outros
		<i>CACTA</i>	Plantas, Metazoários e Fungos
	Crypton	<i>Crypton</i>	Fungos
Classe II Subclasse II	Helitron	<i>Helitron</i>	Plantas, Metazoários e Fungos
	Maverick	<i>Maverick</i>	Metazoários, Fungos e outros

3. REFERÊNCIAS BIBLIOGRÁFICAS

Andrade WM, Ramos MA, Souto WMS, Bento-Silva JS, Albuquerque UP e Araújo EL (2015) Knowledge, uses and practices of the licuri palm (*Syagrus coronata* (Mart.) Becc.) around protected areas in northeastern Brazil holding the endangered species Lear's Macaw (*Anodorhynchus leari*). **Tropical Conservation Science** 8: 893-911.

Angiosperm Phylogeny Group (APG) (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. **Botanical Journal of the Linnean Society** 181: 1-20.

Arnold ML e Kunte K (2017) Adaptive genetic exchange: a tangled history of admixture and evolutionary innovation. **Trends in Ecology & Evolution** 32: 601-611.

Aroucha EPTL e Aroucha ML (2013) **Boas práticas de manejo para o extrativismo sustentável do licuri**. Instituto Sociedade, População e Natureza, Brasília, 92p.

Áquila R, Silva J, Silva JV e Almeida C (2018) The first complete chloroplast genome sequence of *Syagrus coronata*: comparative analysis in Arecoideae (Arecaceae). **Brazilian Journal of Botany** 41: 393-401.

Badaeva ED, Friebe B e Gill BS (1996) Genome differentiation in *Aegilops*. 2. Physical mapping of 5S and 18S-26S ribosomal RNA gene families in diploid species. **Genome** 39: 1150-1158.

Barbosa SN e Almeida C (2018) Genome survey and development of 15 microsatellite molecular markers in *Syagrus coronata* (Martius) Beccari (Arecaceae) by next-generation sequencing. **Brazilian Journal of Botany** 42: 195-200.

Belyayev A, Josefiová J, Jandová M, Kalendar R, Krak K e Mandák B (2019) Natural history of a satellite DNA family: from the ancestral genome component to species-specific sequences, concerted and non-concerted evolution. **Molecular sciences** 20: 1-16.

Belyayev AA e Raskina OM (2010) Dynamics of highly repetitive DNA fraction as indicator of speciation in species of the family Poaceae. **Russian Journal Genetics** 46: 1122-1124.

Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.

Bennetzen JL (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. **Genetica** **115**: 29-36.

Bennetzen JL, MA J e Devos KM (2005) Mechanisms of recente genome size variation in flowering plants. **Annals of Botany** **95**: 127-132.

Bilinski P, Albert PS, Berg JJ, Birchler JA, Grote MN, Lorant A, Quezada J, Swarts K, Yang J e Ross-Ibarra J (2018) Parallel altitudinal clines reveal trends in adaptive evolution of genome size in *Zea mays*. **PLOS Genetics** **14**: 1-19.

Biscotti MA, Olmo E e Heslop-Harrison JS (Pat) (2015) Repetitive DNA in eukaryotic elements. **Chromosome Research** **23**: 415-420.

Blommaert J (2020) Genome size evolution: towards new model systems for old questions. **Proceedings of the Royal Society B: Biological Sciences** **287**: 1-9.

Bonnevier J, Hammerbeck C e Goetz C (2018) Flow cytometry: definition, history and uses in biological research. In: Goetz C, Hammerbeck C and Bonnevier J (eds) **Flow cytometry basics for the non-expert. Techniques in Life Science and Biomedicine for the Non-Expert**. Springer, p. 1-11.

Borém A e Miranda GV (2009) **Melhoramento de Plantas**. UFV, Viçosa, 229p.

Brasil – Ministério de Agricultura, Pecuária e Abastecimento (2014) **Licuri – boas práticas para o extrativismo sustentável orgânico**. MAPA/ACS, Brasília, 49p.

Britten e Kohne (1968) Repeated sequences in DNA. **Science** **161**: 529-540.

Campos JLA, Araújo EL, Gaoue OG e Albuquerque UP (2018) How can local representations of changes of the availability in natural resources assist in targeting conservation?. **Science of the Total Environment** **628-629**: 642-649.

Carvalho AJA, Ferreira MHS e Alves JS (2016) **Manual do licuri**. Áttema Editorial, Salvador, 100p.

Carvalho NOS, Pelacani CR, Rodrigues MOS e Crepaldi IC (2005) Uso de substância reguladoras e não-específicas na germinação de sementes de licuri (*Syagrus coronata* (Mart.) Becc.). **Sitientibus Série Ciências Biológicas** **5**: 28-32.

Carvalho NOS, Pelacani CR, Rodrigues MOS e Crepaldi IC (2006) Crescimento inicial de plantas de licuri (*Syagrus coronata* (Mart.) Becc.) em diferentes níveis de luminosidade. **Revista Árvore** **30**: 351-357.

Castro RA, Fabricante JR e Siqueira Filho JA (2016) A importância da palmeira *Syagrus coronata* (Mart.) Becc. para a conservação da riqueza e diversidade de espécies epífitas vasculares na caatinga. **Revista Árvore** **40**: 1-12.

Chénais B, Caruso A, Hiard S e Casse N (2012) The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. **Gene** **509**: 7-15.

Conceição ACR, Leão MA e Carvalho RF (2018) *Syagrus coronata*. In: Coradin L, Camillo J and Pareyn FGC (eds) **Espécies nativas da flora brasileira de valor econômico atual ou potencial – plantas para o futuro – região Nordeste**. Ministério do Meio Ambiente, Brasília, p. 462-470.

Crepaldi IC, Almeida-Muradian LB, Rios MDG, Penteadó MVC e Salatino A (2001) Composição nutricional do fruto de licuri (*Syagrus coronata* (Martius) Beccari). **Revista Brasileira de Botânica** **24**: 155-159.

Crepaldi IC, Negri G, Salatino A e Costa Neto A (2010) *Syagrus coronata* (licuri) e *Syagrus vagans* (licurioba) (Arecaceae): fibras e ceras foliares de plantas de duas regiões da Bahia. **Sitientibus Série Ciências Biológicas** **10**: 217-221.

Doolittle WF e Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. **Nature** **284**: 601-603.

Douet J e Tourmente S (2007) Transcription of the 5S rRNA heterochromatic genes is epigenetically controlled in *Arabidopsis thaliana* and *Xenopus leavis*. **Heredity** **99**: 5-13.

Dransfield J, Uhl NW, Asmussen CB, Baker WJ, Harley MM e Lewis CE (2005) A new phylogenetic classification of the palm family, Arecaceae. **Kew Bulletin** **60**: 559-569.

Drumond MA (2007) **Licuri *Syagrus coronata* (Mart.) Becc.** Embrapa Semiárido (Embrapa – PE. Documentos, 199), Petrolina, 19p.

Duran C, Appleby N, Edwards D e Batley J (2009) Molecular genetic markers: discovery, applications, data storage and visualisation. **Current Bioinformatics** **4**: 16-27.

Eickbush TH e Eickbush DG (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. **Genetics** **175**: 477-485.

Excoffier L, Smouse PE e Quattro JM (1992) Analysis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. **Genetics** **131**: 479-491.

Fernandes MF e Queiroz LP (2015) Floristic surveys of Restinga Forests in southern Bahia, Brazil, reveal the effects of geography on community composition. **Rodriguésia** **66**: 51-73.

Ferreira ME e Gratapaglia D (1998) **Introdução ao uso de marcadores moleculares em análise genética**. Embrapa Cenargen, Brasília, 220p.

Feschotte C (2008) The contribution of transposable elements to the evolution of regulatory networks. **Nature Reviews Genetics** **9**: 397-405.

Frankel OH, Brown AHD e Burdon JJ (1995) **Conservation of plant biodiversity**. Cambridge University Press, Reino Unido, 2990p.

Freitas Neto M, Pereira TNS, Geronimo IGC, Azevedo AON, Ramos SRR e Pereira MG (2016) Coconut genome size determined by flow cytometry: Tall versus Dwarf types. **Genetics and Molecular Biology** **15**: 1-9.

Garrido-Ramos MA (2015) Satellite DNA in plants: more than just rubbish. **Cytogenetic and Genome Research** **146**: 153-170.

Garrido-Ramos MA (2017) Satellite DNA: an evolving topic. **Genes** **8**: 2-41.

Gomes FM, Ribeiro RTM, Batista MEP, Souza EB, Siqueira Filho JA, Portela S, Castro R e Loiola MIB (2018) *Syagrus coronata* - Licuri. In: Coradin L, Camillo J e Pareyn FGC (eds) **Espécies nativas da flora brasileira de valor econômico atual ou potencial – plantas para o futuro – região Nordeste**. Ministério do Meio Ambiente, Brasília, p. 1056-1072.

Greilhuber J, Dolezel J, Lysak MA e Bennett MD (2005) The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. **Annals of Botany** **95**: 255-260.

Gruendler P, Unfried I, Pascher K e Schweizer D (1991) rDNA intergenic region from *Arabidopsis thaliana* – Structural analysis, intraspecific variation and functional implications. **Journal of Molecular Biology** **221**: 1209-1222.

Guerra M (2008) Chromosome numbers in plant cytotaxonomy: concepts and implications. **Chromosome Evolution** **120**: 339-350.

Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.

Hartl DL e Clark AG (2010) **Princípios de Genética de Populações**. Artmed, Porto Alegre, 660p.

Hinegardner R (1976) Evolution of genome size. In: Ayala FJ (ed) **Molecular Evolution**, Sunderland, p. 179-199.

Hodel RGJ, Segovia-Salcedo C, Landis JB, Crowl AA, Sun M, Liu X, Gitzendanner MA, Douglas NA, Gernam-Aubrey CC, Chen S, Soltis DE et al. (2016) The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. **Applications in Plants Sciences** 4: 1600025.

Hollingsworth PM, Graham SW e Little DP (2011) Choosing and using a plant DNA barcode. **PlosOne** 6: 1-13.

Hughes AFS, Lima FG, Lucchese AM, Goés Neto A e Uetanabaro APT (2013) Antimicrobial Activity of *Syagrus coronata* (Martius) Beccari. **Brazilian Archives of Biology and Technology** 56: 269-274.

IBAMA – Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (2006) **Plano de manejo da arara-azul-de-Lear (*Anodorhynchus leari*)**. IBAMA/Coordenação de Proteção de Espécies da Fauna, Brasília, 80p.

Ibrahim MA, Al-Shomrani BM, Alharbi SN, Elliott TA, Alsuabeyl MS, Alqahtani FH e Manee MM (2021) Genome-wide comparative analysis of transposable elements in Palmae genomes. **Frontiers Bioscience** 26: 1119-1131.

Jamhour J e Laroca SA (2004) A community of wild bees (Hym., Apoidea) of Pato Branco (PR-Brazil): diversity, phenology, floral resources, and biogeographical aspects. **Acta Biológica Paranaense** 33: 27-119.

Johnson D (1996) **Palms: Their Conservation and Sustained Utilization**. IUCN/SSC Palm Specialist Group, Cambridge, 116p.

Kazazian Jr HH (2004) Mobile elements: drivers of genome evolution. **Science** 303: 1626-1632.

Kelly LJ, Renny-Byfield S, Pellicer J, Macas J, Novak P, Neumann P, Lysak MA, Day PD, Berger M, Fay MF, Nichols RA, et al. (2015) Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of the DNA removal characterizes extreme expansions in genome size. **New Phytologist** 208: 596-607.

Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.

Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. **Genetica** **115**: 49-63.

Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES e Adams M (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. **PCR Methods and Applications** **3**: 13-22.

Kit S (1961) Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. **Journal of Molecular Biology** **3**: 711-716.

Kupiec M (2014) Biology of telomeres: lessons from budding yeast. **FEMS Microbiology Reviews** **38**: 144-171.

Lantican DV, Strickler SR, Canama AO, Gardoce RR, Mueller LA e Galvez HF (2019) *De Novo* genome sequence assembly of dwarf Coconut (*Cocos nucifera* L. 'Catigan Green Dwarf') provides insights into genomic variation between Coconut types and related ralm specie. **Genome Report** **9**: 2377-2393.

Lee S e Kim N (2014) Transposable elements and genome size variations in plants. **Genomics & Informatics** **12**: 87-97.

Leitch IJ e Bennett MD (2004) Genome downsizing in polyploid plants. **Biological Journal of the Linnean Society** **82**: 651-663.

Leitch IJ, Johnston E, Pellicer J, Hidalgo O e Bennet MD (2019) **Angiosperm DNA C-values database (release 9.0, Apr 2019)** <https://cvalues.science.kew.org/>. Disponível em: <<https://cvalues.science.kew.org/search/angiosperm>>. Acesso em: 03 maio de 2022.

Leitch IJ, Soltis DE, Soltis PS e Bennett MD (2005) Evolution of DNA amounts across land plants (Embryophyta). **Annals of Botany** **95**: 207–217.

Li FW e Harkess A (2018) A guide to sequence your favorite plant genomes. **Applications in Plant Sciences** **6**: e1030.

Li SF, Su T, Cheng GQ, Wang BX, Li X, Deng CL e Gao WJ (2017) Chromosome evolution in connection with repetitive sequences and epigenetics in plants. **Genes** **8**: 2-19.

Lima DM, Tenório S e Gomes K (2014) Dieta por *Anodorhynchus leari* Bonaparte, 1856 (Aves: Psittacidae) em palmeira de licuri na caatinga baiana. **Atualidades Ornitológicas** **178**: 50-54.

Lin KW e Yan J (2008) Endings in the middle: current knowledge of interstitial telomeric sequences. **Mutation Research** **658**: 95-110.

Lisch D (2013) How important are transposons for plant evolution? **Nature Reviews** **14**: 49-61.

Lorenzi H (2010) **Flora brasileira: Arecaceae (Palmeiras)**. Editora Plantarum, Nova Odessa, 384p.

Lorenzi H, Souza HM, Cerqueira LSC, Costa JTM e Ferreira E (2004) **Palmeiras brasileiras e exóticas cultivadas**. Editora Plantarum, Nova Odessa, 416p.

Loureiro J, Rodriguez E, Dolezel J e Santos C (2007) Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. **Annals of Botany** **100**: 875-888.

Manfio CE, Motoike SY, Santos CEM, Pimentel LD, Queiroz V e Sato AY (2011) Repetibilidade em características biométricas do fruto de macaúba. **Ciência Rural** **41**: 70-76.

McClintock B (1950) The origin and behavior of mutable loci in maize. **Proceedings of the National Academy of Sciences** **36**: 344-355.

Meerow AW, Noblick L, Borrone JW, Covreur TLP, Mauro-Herrera M, Hahn WJ, Kuhn DN, Nakamura K, Oleas NH e Schell RJ (2009) Phylogenetic analysis of seven WRKY genes across the palm Subtribe Attaleinae (Arecaceae) identifies *Syagrus* as Sister Group of the Coconut. **PLoS One** **4**: 1-17.

Mehrotra S e Goyal V (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. **Genomics Proteomics Bioinformatics** **12**: 164-171.

Mestrovic N, Mravinac B, Pavlek M, Vojvoda-Zeljko T, Satovic E e Plohl M (2015) Structural and functional liaisons between transposable elements and satellite DNAs. **Chromosome Research** **23**: 583-596.

Michaelson MJ, Price HJ, Ellison JR e Jhonston JS (1991) Comparison of plant DNA contents determined by Felgen microspectrophotometry and laser flow cytometry. **American Journal of Botany** **78**: 183-188.

Missaggia A e Grattapaglia D (2006) Plant microsatellite genotyping with 4-color fluorescent detection using multiple-tailed primers. **Genetics and Molecular Research 5**: 72-78.

Moraes AP, Lemos RR, Brasileiro-Vidal AC, Soares-Filho WS e Guerra M (2007) Chromosomal markers distinguish hybrids and non-hybrid accessions of mandarin. **Cytogenetic and Genome Research 119**: 275-281.

Morgil H, Gercek YC e Tulum I (2020) Single nucleotide polymorphisms (SNPs) in plant genetics and breeding. In: Çalışkan M, Erol O e Öz GC (eds) **The recent topics in genetic polymorphisms**. IntechOpen, London, 12p.

Moura RF, Dowson DA e Nogueira DM (2017) The use of microsatellite markers in neotropical studies of wild birds: a literature review. **Anais da Academia Brasileira de Ciências 89**: 145-154.

Nass LL, Nishikawa MAN, Fávero AP e Lopes MA (2008) Pré-melhoramento do germoplasma vegetal. In: Nass LL (eds) **Recursos genéticos vegetais**. Embrapa Recursos Genéticos e Biotecnologia, Brasília, p. 683-716.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. **Genetics 89**: 583-590.

Neumann P, Novak P, Hostakova N e Macas J (2019) Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. **Mobile DNA 10**: 2-17.

Neves, Geís Ferreira. **Caracterização de populações naturais de licuri, *Syagrus coronata* (Mart.) Becc. (Arecaceae): pré-melhoramento e conservação da espécie**. 2021. 80f. Dissertação (Mestrado em Fitotecnia) – Universidade Federal de Viçosa, Viçosa, 2021.

Neves N, Delgado M, Silva M, Caperta A, Morais-Cecílio L e Viegas W (2005) Ribosomal DNA heterochromatin in plants. **Cytogenetic and Genome Research 109**: 104-111.

Noblick LR (1986) Palmeiras das caatingas da Bahia e as potencialidades econômicas. In **Simpósio sobre a caatinga e sua exploração regional**, Embrapa-DDT, Feira de Santana, p. 99-115.

Noblick LR (2017) A revision the genus *Syagrus* (Arecaceae). **Phytotaxa 294**: 1-262.

Noblick LR e Barreto RC (2018) *Syagrus coronata* - Licuri. In: Coradin L, Camillo J and Pareyn FGC (eds) **Espécies nativas da flora brasileira de valor econômico atual ou potencial – plantas para o futuro – região Nordeste**. Ministério do Meio Ambiente, Brasília, p. 1206-1211.

Normas Brasil (2008) **Instrução Normativa IBAMA nº 191 de 24/09/2008**. Portal de normas e legislação brasileira. Disponível em: http://www.normasbrasil.com.br/norma/instrucao-normativa191/2008_75905.html. Acesso em 14 de agosto, 2021.

Novak P, Neumann P e Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. **BCM Bioinformatics 11**: 2-12.

Novak P, Robledillo LA, Koblizková A, Vrbová I, Neumann P e Macas J (2017) TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. **Nucleic Acids Research 45**: e111.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. **Molecular Ecology 13**: 1143-1155.

Oliveira UR, Espírito-Santo FS e Alvarez IA (2015) Comunidade epifítica de *Syagrus coronata* (Mart.) Becc. em áreas de pastagens da Caatinga, Bahia. **Revista Caatinga 28**: 84-91.

Orgel LE e Crick FHC (1980) Selfish DNA: the ultimate parasite. **Nature 284**: 604-607.

Pellicer J, Fay MF e Leitch IJ (2010) The largest eukaryotic genome of them all?. **Botanical Journal of the Linnean Society 164**: 10-15.

Pellicer J, Hidalgo O, Dodsworth S e Leitch J (2018) Genome size diversity and its impact on the evolution of land plants. **Genes 9**: 1-14.

Piégu B, Bire S, Arensburger e Bigot Y (2015) A survey of transposable element classification systems – A call for a fundamental update to meet the challenge of their diversity and complexity. **Molecular Phylogenetics and Evolution 86**: 90-109.

Rebollo R, Romanish MT e Mager DL (2012) Transposable elements: an abundant and natural source of regulatory sequences for host genes. **Annual Review of Genetics 46**: 21-42.

Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.

Richards EJ e Ausubel FM (1988) Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. **Cell** **53**: 127-136.

Roa F e Guerra M (2012) Distribution of 45S rDNA sites in chromosomes of plants: structural and evolutionary implications. **BCM Evolutionary Biology** **12**: 2-13.

Roa F e Guerra M (2015) Non-random distribution of 5S rDNA sites and its association with 45S rDNA in plant chromosomes. **Cytogenetic and Genome Research** **146**: 243-249.

Rocha KMR (2009) Licuri: a palmeira da vida das araras e dos sertanejos. **Revista Ouricuri** **1**: 201-214.

Rocha, Kilma Manso Raimundo. **Biologia reprodutiva da palmeira licuri (*Syagrus coronata* (Mart.) Becc.) (Arecaceae) na ecorregião do Raso da Catatina, Bahia**. 2009.100f. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal Rural de Pernambuco, Recife, 2009.

Rodrigues MOS, Lima-Brito A, Carvalho NOS, Crepaldi IC e Pelacani CR (2006) Viabilidade de sementes de licuri (*Syagrus coronata* (Mart.) Becc.) durante o armazenamento. **Sitientibus Série Ciências Biológicas** **6**: 50-55.

Rufino MUL, Costa JTM, Silva VA e Andrade LHC (2008) Conhecimento e uso do ouricuri (*Syagrus coronata*) e do babaçu (*Orbignya phalerata*) em Buíque, PE, Brasil. **Acta Botanica Brasilica** **22**: 155-159.

Salles KTSL, Meneghetti SMP, Salles WFL, Meneghetti MR, Santos ICF, Silva JPV, Carvalho SHV e Soletti JI (2010) Characterization of *Syagrus coronata* (Mart.) Becc. oil and properties of methylesters for use as biodiesel. **Industrial Crops and Products** **32**: 518-521.

Santana JP, Rocha PA, Silva TR, Ribeiro AS e Prata APN (2017) Floristic characterization of Ibura national forest, Sergipe, Brazil. **Bioscience Journal** **33**: 447-464.

Santos-Moura SS, Gonçalves EP, Melo LDFA, Paiva LG e Silva TM (2016) Morphology of fruits, diaspores, seeds, seedlings, and saplings of *Syagrus coronata* (Mart.) Becc. **Bioscience Journal** **32**: 652-660.

Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.

Santos-Neto JR e Camandaroba M (2008) Mapeamentos dos sítios de alimentação da arara-azul-de-Lear (*Anodorhynchus leari*) (Bonaparte, 1856). **Ornithologia 3**: 1-17.

Simpson EH (1949) Measurement of diversity. **Nature 163**: 688.

Schmidt G e Thannhauser (1945) A method for the determination of desoxiribonucleic acid, ribonucleic acid, and phosphoproteins in animal tissues. **Journal of Biological Chemistry 161**: 83-89.

Schubert I e Lysak MA (2011) Interpretation of karyotype evolution should consider chromosome structural constraints. **Trends in Genetics 27**: 207-216.

Schubert I e Vu GTH (2016) Genome stability and evolution: attempting a Holistic view. **Trends in Plant Science 21**: 749-757.

Semagn K, Bjornstard e Ndjioudjop MN (2006) An overview of molecular marker methods for plants. **African Journal of Biotechnology 5**: 2540-2568.

Stoddart JA e Taylor JF (1987) Genotypic diversity: estimation and prediction in samples. **Genetics 118**: 705-711.

Souza G, Vanzela ALL, Crosa O e Guerra M (2016) Interstitial telomeric sites and Robertsonian translocations in species of *Ipheidion* and *Nothoscordum* (Amaryllidaceae). **Genetica 144**: 157-166.

Souza MCP, Moura F, Silva JV e Almeida C (2018) Phylogeography of the palm *Syagrus coronata* (Martius) Beccari (Arecaceae): distribution in the “Caatinga” and Atlantic Forest domains. **Brazilian Journal of Botany 41**: 849-857.

Souza MM, Palomino G, Pereira TNS, Pereira MG e Viana AP (2004) Flow cytometric analysis of genome size variation in some *Passiflora* species. **Hereditas 141**: 31-38.

Shannon CE (2001) A mathematical theory of communication. **Mobile Computing and Communications review 5**: 3-55.

Shinozaki K, Ohme M, Tanaka M, Wakasuki T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, et al. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. **The Embo Journal 5**: 2043-2049.

Storchová H e Olson MS (2007) The architecture of the chloroplast *trnH-psbA* noncoding region in angiosperms. **Plant Systematics and Evolution** **268**: 235-256.

Thomson MJ (2014) High-throughput SNP genotyping to accelerate crop improvement. **Plant Breeding and Biotechnology** **2**: 195-212.

Vaio M, Gardner A, Speranza P, Emshwiller E e Guerra M (2016) Phylogenetic and cytogenetic relationships among species of *Oxalis* section *Articulatae* (Oxalidaceae). **Plant Systematics and Evolution** **302**: 1253-1265.

Volkov RA, Medina FJ, Zentgraf U e Hemleben V (2004) Molecular cell biology: organization and molecular evolution of rDNA, nucleolar dominance, and nucleolus structure. In: Esser K, Luttge U, Beyschlag W e Murata J (eds) **Molecular cell biology**. Progress, p. 106-146.

Xiao Y, Xu P, Fan H, Baudouin L, Xia W, Bocs S, Xu J, Li Q, Guo A, Zhou L, Li J, Wu Y, Ma Z, Armero A, Issali AE, Liu N, Peng N e Yang Y (2017) The genome draft of cocnut (*Cocos nucifera*). **Gigascience** **6**: 1-11.

Weiss-Schneewiss H, Leitch AR, McCann J, Jang TS e Macas J (2015) Employing next generation sequencing to explore the repeat landscape of the genome. In: Horandl E, Appelhans MS (eds) **Next-generation sequencing in plant systematics**. Scientific Books, Königstein, p. 1-25.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. (2007) A unified classification system for eukaryotic transposable elements. **Perspectives** **8**: 973-982.

Wright S (1978) **Evolution and genetics of populations**. University of Chicago Press, Chicago, 511p.

Yang Y, Bocs S, Fan H, Armero A, Baudouin L, Xu P, Xu J, This D, Hamelin C, Iqbal A, et al. (2021) Coconut genome assembly enables evolutionary analysis of palms and highlights signaling pathways involved in salt tolerance. **Communications Biology** **4**: 105.

Zhu Q, Cai Z, Tang Q e Jin W (2016) Repetitive sequence analysis and karyotyping reveal different genome evolution and speciation of diploid and tetraploid *Tripsacum dactyloides*. **The Crop Journal** **4**: 247-255.

CAPÍTULO II

Genome size and *de novo* identification repetitive fraction in the *Syagrus coronata* (Arecaceae)

Genome size and de novo identification repetitive fraction in the *Syagrus coronata* (Areaceae)

Jackeline Terto da Silva Santana^{1,3}, Merieli Araújo do Carmo Silva¹, Ewerton Cavalcante Massena¹, Luiz Gustavo Rodrigues Souza², André Marques¹, Cícero Almeida¹, Reginaldo de Carvalho³

¹*Laboratório de Recursos Genéticos, Universidade Federal de Alagoas, Campus de Arapiraca, Brasil*

²*Laboratório de Citogenética e Evolução, Universidade Federal de Pernambuco, Brasil*

³*Laboratório de Citogenética, Universidade Federal Rural de Pernambuco, Brasil*

Autor para correspondência:

Cícero de Almeida

Campus Arapiraca

Universidade Federal de Alagoas

Avenida Manoel Severino Barbosa s/n, Rodovia AL 115, km 6,5. Bairro Bom Sucesso

Arapiraca, AL, Brasil

Telefone: (82) 3482-1831

E-mail: cicero@arapiraca.ufal.br

ORCID: <https://orcid.org/0000-0002-6793-8787>

Abstract

Plant genomes range greatly in size and composition, being the repetitive DNA the most important to the diversity, representing up to 90% of the genome. Repetitive DNA is composed of satellite sequences and transposable elements, which are mainly responsible for the differentiation in genomic size. *Syagrus coronata* is an endemic palm tree from Brazil, it belongs to the Arecaceae family, with geographic distribution in the *Caatinga* and Atlantic Forest domains. This study aims to estimate the genome size and to characterize repetitive sequences in *S. coronata*. Genome size was estimated by flow cytometry and *k-mers* approach using method RESPECT. The repetitive sequences were analyzed using paired-end reads using RepeatExplorer pipelines. The genome size was estimated to be 2.27 ± 0.026 Gbp/1C using flow cytometry and 2.76 Gbp using *k-mers* approach. The repetitive analysis showed that repetitive sequences correspond to 60.39% of genome. The LTR retrotransposons were most abundant (49.84%), being the *Ty1/Copia* and *Ty3/Gypsy* making up 37,62% and 12,2%, respectively. Others repetitive elements as transposons, ribosomal DNA and satellites sequences making up 3.91%, 0.56% and 1.32%, respectively. The satellites sequences showed high diversity in number and monomers length, however showed minor contribution in the genome size. We conclude that *S. coronata* to have genome moderately repetitive DNA sequences, being the LTR retrotransposons by far the most abundant.

keywords: ouricuri, palm tree, 1C-value, genome length

Introduction

Genome size has been extensively estimated by flow cytometry, being designed as C-value (total amount of DNA in the haploid nucleus)(Pellicer and Leitch, 2019), which it is necessary for whole-genome sequencing and comparative genomic studies. The size of genomes in angiosperms shows great variation, with a minimum genome size of 65.00 Mbp/1C in *Genlisea tuberosa* Rivadavia, Gonella & A. Fleischm. (Lentibulariaceae family) (Fleischmann et al., 2014) and the gigantic genome *Paris japonica* (Franch. & Sav.) Franch. (Melanthiaceae family) (148,851 Mbp/1C; Pellicer et al., 2018). Areaceae showed a estimate of the minimum genome size in *Korthasis ridiga* (744.80 Mbp/1C; Kumar et al., 2011) and maximum in *Voanioala gerardii* (29,400 Mbp/1C; Johnson et al., 1989). In the *Syagrus* genus, the genome size has been reported for *S. romanzoffiana* (3,038 Mbp/1C) and *S. glaucescens* (3,332 Mbp/1C)(Röser et al., 1997).

Plant genome size variation may be explained by several mechanisms, as the polyploidy events, genome duplication, proliferating transposable elements and tandem repeats (Blommaert, 2020). Among all events, the repetitive fraction of the genome explains the majority of the genome size variation in the angiosperms. Repetitive fraction in plants consists of two categories of repeat types, the repeat pattern dispersed throughout the genome, which include transposable elements (TEs) and the tandem repeats, which include ribosomal DNAs (5S and 35S), satellite DNAs, among others. TEs are reported as fundamental sequences to maintain the integrity of centromeres and telomeres (Kazazian Jr, 2004; Rebollo et al., 2012). They play a role in gene regulation, structural and functional organization of genomes (Kidwell, 2002; Kazazian Jr, 2004; Feschotte, 2008) and have the ability to move and multiply, being widely variable in the sequences and structure. TEs are classified based on their mode of replication, being divided into two classes, retrotransposons (*class I*) and

DNA transposons (*class II*), in which the first class may be subdivided into two subgroups depending on the presence or absence of Long Terminal Repeat (LTR) (Mestrovic et al., 2015).

The tandem repetitive DNAs are organized in long monomer arrays and are frequently found in specific parts on the chromosomes, such as centromeres, telomeres and interstitial regions, being the main components of the constitutive heterochromatin, suggesting contributing to chromosomal integrity and regulation of gene expression. (Mehrotra & Goyal, 2014; Mestrovic et al., 2015). Depending on the length of the monomer, they can be classified into microsatellites, minisatellites and satellite DNA sequences (DNAsat) (Mehrotra & Goyal, 2014). The DNAsat are characterized by representing the most variable fraction of the genome, ranging from 1,000 to 100,000 copies (Mehrotra & Goyal, 2014; Belyayev et al., 2019). Genomic evolution and genome sizes are directly related to repetitive DNA components (Kelly & Leitch, 2011; Mehrotra & Goyal, 2014). Genome size estimates and characterization of the repetitive fraction provide an understanding of genome evolution, allowing to correlate genome size with specific repetitive DNA sequences and with the divergence time between related species.

The Arecaceae family is characterized by contain the tallest domesticated trees and longest-lived species (Hadrami et al., 2012) and are of considerable economic importance in arid and semi-arid regions. Genomic study has been reported in four species (*Phoenix dactylifera*, *Cocos nucifera*, *Calamus simplicifolius* and *Elaeis oleifera*), which the genome-wide annotation of TEs showed that *class I* elements formed the majority of the TEs, making up 36% to 76% (Ibrahim et al., 2021).

Syagrus coronata (Mart.) Becc. belongs to the Arecaceae family, Arecoideae subfamily, Coccoceae tribe (Dransfield et al., 2005). The genus *Syagrus* originated approximately 35 million years ago, has about 30 species and diversification among species

originated 27.04 million years ago, with the genus *Cocos* as the closest (Meerow et al., 2009). *Syagrus coronata* is an endemic species from Brazil, known as “licuri” and “ouricuri”, it is distributed in the *Caatinga* (seasonally dry tropical biome) and *Atlantic Forest domains* (tropical forest biome) (Souza et al., 2018). The species is one of the most important palm trees in Brazil, it has great social, economic and ecological importance, being used both in cooking and in industry, as well as in domestic utensils and food for native fauna, producing fruits in periods of drought (Andrade et al., 2015).

Genetic studies have been carried out in *S. coronata*, such as the development of microsatellite molecular markers (Barbosa and Almeida, 2018) and genetic structuring in the *Caatinga* and Atlantic Forest domains (Souza et al., 2018). Genomic studies are available for the chloroplast (Áquila et al., 2018) and mitochondrial (unpublished data) genomes. However, for the nuclear genome there are no studies that allow an understanding of the genomic evolution of *S. coronata*. To understand the evolution of the *S. corona* genome, this study extend the aforementioned results to test hypotheses: (1) The repetitive DNA composition of the *S. coronata* genome is greatly constituted by retronsposons, and (2) the tandem repeats are remarkable diversity in *S. coronta*. To answer these questions, genome size was estimated using flow cytometry and *k-mers* approach, which is the most reliable and predominantly used method in plants and the composition of repetitive elements was characterized using the RepeatExplorer pipeline.

Materials and Methods

Plant material, DNA Extraction and Sequencing

Plant material was collected from adult palm tree in the state of Alagoas, Brazil (9°45'09"S; 36°39'40"W), and approximately 2 cm² of leaves was used to extract the genetic material (including nuclear, chloroplast, and mitochondrial DNA), following the Cetyl

Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.

Trimethylammonium Bromide - CTAB extraction method (Doyle and Doyle, 1987). DNA integrity was analyzed on a 1% agarose gel and quantified using a spectrophotometer.

The sequencing was performed at the Central Laboratory of High-Performance Technologies in Life Sciences (LacTad) at the State University of Campinas – UNICAMP, São Paulo, Brazil where DNA molecules were fragmented into approximately 200-500 bp to build the sequencing libraries. Then, the fragments were ligated with adapters using the “Nextera DNA Sample Preparation” kit (Illumina) and the paired-end reads with a size of 100 nt were sequenced using the Illumina HiSeq2500 platform.

Estimation of genome size

For flow cytometry, a suspension of nuclei from young leaf tissue was prepared as described by Loureiro et al. (2007) using WPB buffer. Genome size was estimated using a CyFlow SL flow cytometer (Partec, Görlitz, Germany). The final DNA content for each adhesion was calculated based on at least three different measurements. As an internal control, young leaf tissue of the maize cultivar (*Zea mays* L.) CE-777 (2C = 5.43 pg) (Lysák & Doležel., 1998) was used. FloMax software (Partec) was used for data processing. The genome size was estimated by counting *k-mers* ($k = 31$) in the software Jellyfish-2 v.2.3.0 (Marçais and Edwards, 2011) and *k-mer* frequency distribution was processed using the REPEAT SPECTra Estimation (RESPECT) v.1.3.0 (Sarmashghi et al., 2021).

Graph-based cluster of Illumina reads using RepeatExplorer

The sequences were submitted to qualification analysis in the FastQC statistical program and the reads from the chloroplast and mitochondrial genomes were removed using Geneious R.91 software. A total of 2,601,514 reads were used to obtain the clusters using the RepeatExplorer pipeline, using the minimum overlap (K-mer) parameters of 40 bp; minimum

length of 55 nt to be considered significant for the formation of each cluster and clusters with at least 0.01% repetition in the genome were characterized. An analysis of the contribution of repetitive DNA to genome size was performed in 101 species, distributed in 37 Orders (Table S2), with data published in several studies (Vitales et al., 2020; Zwyrková et al., 2020; Santos et al., 2020; Wang et al., 2020).

Results

Genome size

The C-value obtained for the genome size estimate using flow cytometry of *S. coronata* was 2C-Value 4.64 ± 0.05 pg, which corresponds to 2.27 ± 0.026 Gbp/1C, being considered a small genome (Figure 1). The haploide genome size estimated using a *k-mer* approach was 2.76 Gbp, with coverage 5.09x.

Repetitive analysis

The RepeatExplorer analysis showed 81,820 clusters, of which 227 clusters correspond to more than 0.01% of genome repetition. The repetitive fraction corresponded to 60.39% of the genome and among the repetitive elements, the retrotransposons make up major the repetitive DNA, being the LTR retrotransposons most abundant (49.84%) (Table 1). Among the LTR, 37.62% were identified as belonging to the Ty1/Copia superfamily and 12.22% as belonging to the Ty3/Gypsy superfamily. Within the Ty1/Copia superfamily it was possible to identify four families, as the Angela the most most abundant, corresponding to 26.28% of the genome. In the Ty3/Gypsy superfamily, four families were identified, where Ogre/Tat constituted the most abundant with a proportion of 9.27% of the genome (Table 1). The TEs transposons showed to be 3.92%, distributed in three families, with emphasis on Em/Spm/CACTA which represented 3.53%.

Ribosomal DNAs (rDNA) corresponded to 0.56% and satellite DNA make up 1.32% of the genome. Eleven satellite DNA families (called SatSCO1-SatSCO11) were observed, of which the monomer sizes ranged from 21 to 276 bp, distributed into five satellites (monomers larger than 100 bp) and six minisatellites (monomers ranging from 20-100 bp) (Table S1). The repetitive sequences showed variation for CG content, which ranged from 23% to 69.6%. Among the satellites, the SatSCO02 satellite (110 bp monomer) the most abundant (0.26%), while the most abundant minisatellite was the SatCO1 (55 bp monomer) with 0.76% (Table 2).

Correlation between repetitive DNA and genome size

The proportion of repetitive elements was correlated with genome size, retrotransposons, transposons and tandem repeats (rDNA and satellite sequences). The repetitive sequences are correlated with genome size, indicating to be main component of the genome (Figure 3A). The repetitive proportion was highly correlated with retrotransposons (Figure 3 B), while transposons and tandem repeats showed minor contribution (Figure 3C and 3D).

Discussion

Genome size

Genome size refers to the amount of DNA in the cell nucleus and usually is expressed as the C-value (Pellicer et al., 2018). Here, we estimated the genome size in *S. coronata* ($2n = 32$; Moro et al., 1999) using two methods: flow cytometry and *k-mers* approach and the two methods showed similar results. The genome size of *S. coronata* is close to other Arecoideae species, such *Cocos nucifera* ($2n = 32$) that revealed a genomic size of 2.42 Gpb (Xiao et al., 2017; Yang et al., 2021) and 2.23 Gpb (Lantican et al., 2019, estimated by kmer approach),

and 2C-Value=5.72 - 5.48 pg, estimated by flow cytometry (Freitas Neto et al., 2016). According to classification Leitch et al. (2005), the *S. coronata* and *C. nucifera* have small genomes (1C-Value \leq 3.5 pg). The genome size variation in plants is due multiple factors such as polyploidy and abundance of repetitive DNA, which may make up to 90% of the genome of plant species (Mehrotra & Goyal, 2014; Pellicer et al. , 2018; Belyayev et al., 2019).

Repetitive sequences

We investigated the abundance of DNA repetitive sequences in *S. coronata* by graph-based clustering using data from next-generation sequencing (Novák et al., 2013). This study also characterized of repetitive fraction in the *S. coronata* genome, being revealed that DNA repetitive sequences comprised 60.39% of the genome and for *Cocos nucifera*, 72.75% of the genome was characterized by repetitive DNA, being a strong predominance of TEs (92.23%) (Yang et al., 2021), while in *S. coronata* the TEs constituted 89% of the genome, suggesting that Arecoideae subfamily shows high proportion of TEs. Other study in Arecaceae showed that repetitive sequence ranged from 41.42% (*P. dactylifera*) to 81.25% (*C. nucifera*) (Ibrahim et al., 2021).

In other orders, as the comparison in 37 different Orders (101 species), showed that the abundance of DNA repetitive sequences is variable between Orders and between species of the same Order (Figure 2, Table S2). The correlation between genome size and abundance of repetitive sequences showed positive, however some species showed big genome size and small amount repetitive sequences (Figure 3A), possibly polyploidy events to be responsible for increasing the genome size. The retrotransposons show a high correlation with the amount of repetitive DNA (Figure 3B), suggesting that they contribute for genome expansion, possibly due to the ability to change its position (Feschotte et al., 2002; Feschotte, 2008).

Syagrus coronata showed similar pattern to other species, which the retrotransposons were the main component of the genome.

Studies on the constitution of the repetitive fraction in members of the Poales Order, such as *Zea mays* L. (Meyers et al., 2001) and *Triticum aestivum* L. (Brenchley et al., 2012), are well established and show that more than 50% of the genome is constituted by retrotransposons, being characterized as fundamental components in the composition of the genome of these species. Considering that the Order Arecales is very close to the Poales and taking into account that the repetitive sequences in the genome of the species of both orders have similarity in the composition of the sequences, it is suggested that part of this inheritance occurred before the separation of the Orders, evidencing that the Evolution of repetitive DNA of species belonging to Arecales follows the same pattern of species of order Poales.

On the other hand, the proportion of transposons identified in *S. coronata* was significantly lower than the retrotransposons, showing no relative importance in the increase of the *S. coronata* genome. In plants, the contribution of transposons to genome sizes does not follow a significant correlation (Figure 3C).

Satellite DNA

Satellite DNA sequences in *S. coronata* show a high diversity in number, types and size of monomers. However, it represents only 1.88% of the total genome, suggesting that it does not have a high contribution to the increase in genome size. On the other hand, it revealed that the genome of *S. coronata* is quite diversified in terms of satellite sequences, which may be associated with constitutive heterochromatin and that future cytomolecular studies will allow us to understand the chromosomal evolution of the species. The contribution of satellite DNA to the genome size of plant species is variable, in which some

species present large contributions and in others the contribution is of lesser relevance, not being a rule for the different species analyzed (Figure 3D).

It is concluded that the retroelements seem to be responsible for an important part of the contribution in the increase of the genome size of *S. coronata* and possible influence on the structural and functional dynamics of the current karyotype. The next step of the research will be the construction of probes, based on the repetitive sequences of the LTRs, and to verify the pattern of bands and the distribution of these elements along the chromosomal complement. These results can be applied in other species of the genus *Syagrus*, using the same repetitive DNA probes for a comparative study of the occurrence and distribution of LTRs and rDNA in the karyotypes of *S. coronata* and related species. In addition, the probes may assist in the characterization of accessions in germplasm banks.

Acknowledgment

We thank the Universidade Federal de Alagoas (UFAL) for the laboratories used and scientific support and the Fundação de Amparo à Pesquisa do Estado de Alagoas (FAPEAL) for funding this project.

Interest conflicts

The authors declare that there are no conflicts of interest in this research.

Author contribution JT, MACS and EM development of methodology, JT, GS, RC, AM conception and design and CA writing, review, and revision of the manuscript. All authors read and approved the final version.

References

Andrade WM, Ramos MA, Souto WMS, Bento-Silva JS, Albuquerque UP and Araújo EL (2015) Knowledge, uses and practices of the licuri palm (*Syagrus coronata* (Mart.)

Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.

Becc.) around protected areas in northeastern Brazil holding the endangered species Lear's Macaw (*Anodorhynchus leari*). Trop Conserv Sci 8:893-911.

Áquila R, Silva J, Silva JV and Almeida C (2018) The first complete chloroplast genome sequence of *Syagrus coronata*: comparative analysis in Arecoideae (Arecaceae). Rev Bras Bot 41: 393-401.

Barbosa SN and Almeida C (2018) Genome survey and development of 15 microsatellite molecular markers in *Syagrus coronata* (Martius) Beccari (Arecaceae) by next-generation sequencing. Rev Bras Bot 41:1-6.

Belyayev A, Josefiová J, Jandová M, Kalendar R, Krak K and Mandak B (2019) Natural history of a Satellite DNA family: from the ancestral genome component to species-specific sequences, concerted and non-concerted evolution. Int J Mol Sc 20:1-16.

Blommaert J (2020) Genome size evolution: towards new model systems for old questions. Proc R Soc B 287: 1-9.

Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, *et al.* (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705-710.

Doyle JJ and Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-15.

Dransfield J, Uhl NW, Asmussen CB, Baker WJ, Harley MH and Lewis CE (2005) A new phylogenetic classification of the palm family, Arecaceae. Kew Bull 60:559-569.

Feschotte C (2008) The contribution of transposable elements to the evolution of regulatory networks. Nat Rev Genet 9:397-405.

Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.

Feschotte C, Jiang N and Wessler SR (2002) Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3:329-341.

Fleischmann A, Michael TP, Rivadavia F, Sousa A, Wang W, Temsch EM, Greilhuber J, Müller KF, Heubl G (2014) Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. *Ann Bot* 114:1651-1663.

Freitas Neto M, Pereira TNS, Geronimo IGC, Azevedo AON, Ramos SRR and Pereira MG (2016) Coconut genome size determined by flow cytometry: Tall versus Dwarf types. *Genet Mol Biol* 15:1-9.

Hadrami A, Al-Khayri JM (2012) Socioeconomic and traditional importance of date palm. *Emirates Journal of Food and Agriculture* 24: 371.

Ibrahim MA, Al-Shomrani BM, Alharbi SN, Elliott TA, Alsuabeyl MS, Alqahtani FH, Manee MM (2021) Genome-wide comparative analysis of transposable elements in *Palmae* genomes. *Front Biosci* 26:1119-1131.

Johnson MAT, Kenton AY, Bennett MD, Brandham PE (1989) *Voanioala gerardii* has the highest known chromosome number in the monocotyledons. *Genome* 32: 328-333.

Kazazian Jr HH (2004) Mobile elements: drivers of genome evolution. *Science* 303:1626-1632.

Kelly LJ and Leitch IJ (2011) Exploring giant plant genomes with next-generation sequencing technology. *Chromosome Res* 19:939-953.

Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115:49-63.

- Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.
- Kumar P, Turner I, Nagaraja Rao A, Arumuganathan K (2011) Estimation of nuclear DNA content of various bamboo and rattan species. *Plant Biotechnology Reports* 5: 317-322.
- Lantican DV, Strickler SR, Canama AO, Gardoce RR, Mueller LA and Galvez HF (2019) De Novo genome sequence assembly of dwarf Coconut (*Cocos nucifera* L. ‘Catigan Green Dwarf’) provides insights into genomic variation between Coconut types and related ralm specie. *Gen Rep* 9: 2377-2393.
- Leitch IJ, Soltis DE, Soltis PS and Bennett MD (2005) Evolution of DNA amounts across land plants (Embryophyta). *Ann Bot* 95:207–217.
- Loureiro J, Rodriguez E, Dolezel J and Santos C (2007) The new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Ann Bot* 100:875-888.
- Lysak MA and Dolezel J (1998) Estimation of nuclear DNA content in *Sesleria* (Poaceae). *Caryologia* 51:123-132.
- Meerow AW, Noblick L, Borrone JW, Covreur TLP, Mauro-Herrera M, Hahn WJ, Kuhn DN, Nakamura K, Oleas NH and Schell RJ (2009) Phylogenetic analysis of seven WRKY genes across the palm Subtribe Attaleinae (Arecaceae) identifies *Syagrus* as Sister Group of the Coconut. *PLoS One* 4:1-17.
- Mehrotra S and Goyal V (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genomics Proteomics Bioinformatics* 12:164-171.
- Meštrović N, Mravinac B, Pavlek M, Vojvoda-Zeljko T, Šatović E and Plohl M (2015) Structural and functional liaisons between transposable elements and satellite DNAs. *Chromosome Res* 23:583-596.

- Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.
- Meyers BC, Tingey SV and Morgante M (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res* 11:1660-1676.
- Moro JR Silva MAS and Geraldo JS (1999) Kariological study of five Arecoideae palmae species. *Acta Horti* 486: 99-102.
- Novák P, Neumann P, Pech J, Steinhaisl J and Macas J (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29:792-793.
- Pellicer and Leitch IJ (2019) The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytologist* 226:301-305.
- Pellicer J, Hidalgo O, Dodsworth S and Leitch IJ (2018) Genome size diversity and its impact on the evolution of land plants. *Genes* 9:1-14.
- Rebollo R, Romanish MT and Mager DL (2012) Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet* 46:21-42.
- Röser M, Johnson MAT, Hanson L (1997) Nuclear DNA amounts in palms (Arecaceae). *Botanica Acta* 110: 79-89.
- Santos V, Silva EF and Almeida C (2020) Genome size and identification of repetitive DNA sequences using low coverage sequencing in *Hancornia speciosa* Gomes (Apocynaceae: Gentianales). *Genet Mol Bio* 43: 1-4.
- Sarmashghi S, Balaban M, Rachtman E, Touti B, Mirarab S, Bafna V (2021) Estimating repeat spectra and genome length from low-coverage genome skims with RESPECT. *Plos Computational Biology*. <https://doi.org/10.1371/journal.pcbi.1009449>

Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.

Souza MCP, Moura F, Silva JV and Almeida C (2018) Phylogeography of the palm *Syagrus coronata* (Martius) Beccari (Arecaceae): distribution in the “Caatinga” and Atlantic Forest domains. *Rev Bras Bot* 41:849-857.

Vitales D, Álvarez I, Garcia S, Hidalgo O, Feliner GN, Pellicer J, Vallès J and Garnatje T (2020) Genome size variation at constant chromosome number is not correlated with repetitive DNA dynamism in *Anacyclus* (Asteraceae). *Annals of Botany* 125: 611-623.

Wang D, Zheng Z, Li Y, Hu H, Wang Z, Du X, Zhang S, Zhu M, Dong L, Ren G and Yang Y (2020) Which factors contribute most to genome size variation within angiosperms? *Ecol Evol* 2021: 2660-2668.

Xiao Y, Xu P, Fan H, Baudouin L, Xia W, Bocs S, Xu J, Li Q, Guo A, Zhou L, Li J, Wu Y, Ma Z, Armero A, Issali AE, Liu N, Peng N and Yang Y (2017) The genome draft of coconut (*Cocos nucifera*). *Gigascience* 6:1-11.

Yang Y, Bocs S, Fan H, Armero A, Baudouin L, Xu P, Xu J, This D, Hamelin C, Iqbal A, *et al.* (2021) Coconut genome assembly enables evolutionary analysis of palms and highlights signaling pathways involved in salt tolerance. *Commun Biol* 4:105.

Zwyrtková J, Němečková A, Čížková J, Holušová K, Kapustová V, Svačina R, Kopecký D, Till BJ, Doležel J and Hříbová E (2020) Comparative analyses of DNA repeats and identification of a novel Fesreba centromeric element in fescues and ryegrasses. *BMC Plant Genome* 20: 1-15.

Internet Resources

Angiosperm DNA C-values database, <https://cvalues.science.kew.org/> (January 11, 2022).

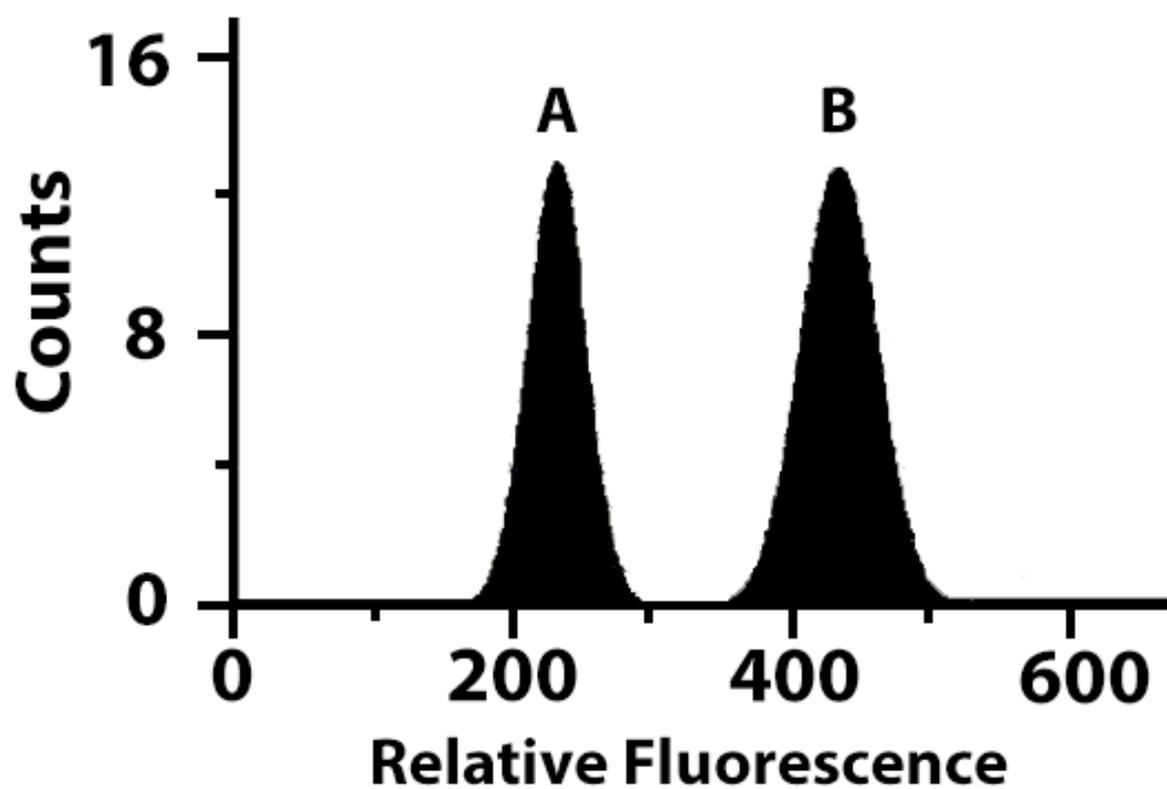


Figure 1. Flow cytometric histogram of relative fluorescence intensity of *Syagrus coronata* (A) and *Zea mays* CE-777 (B).

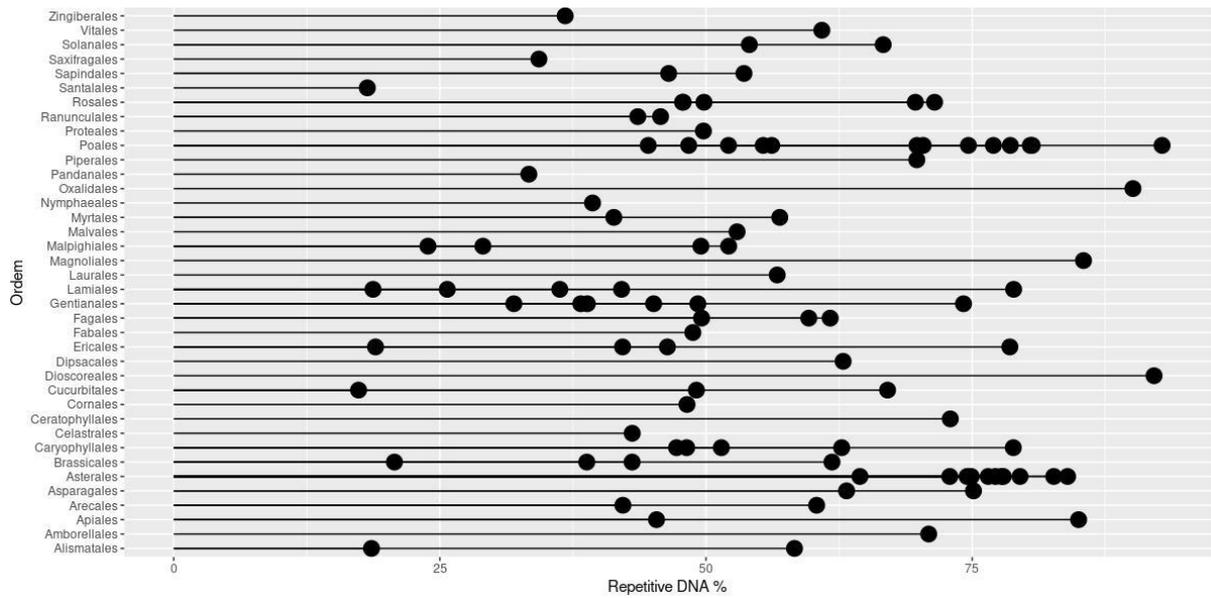


Figure 2. Repetitive DNA fraction in the genome of 101 plants, distributed in 37 Orders.

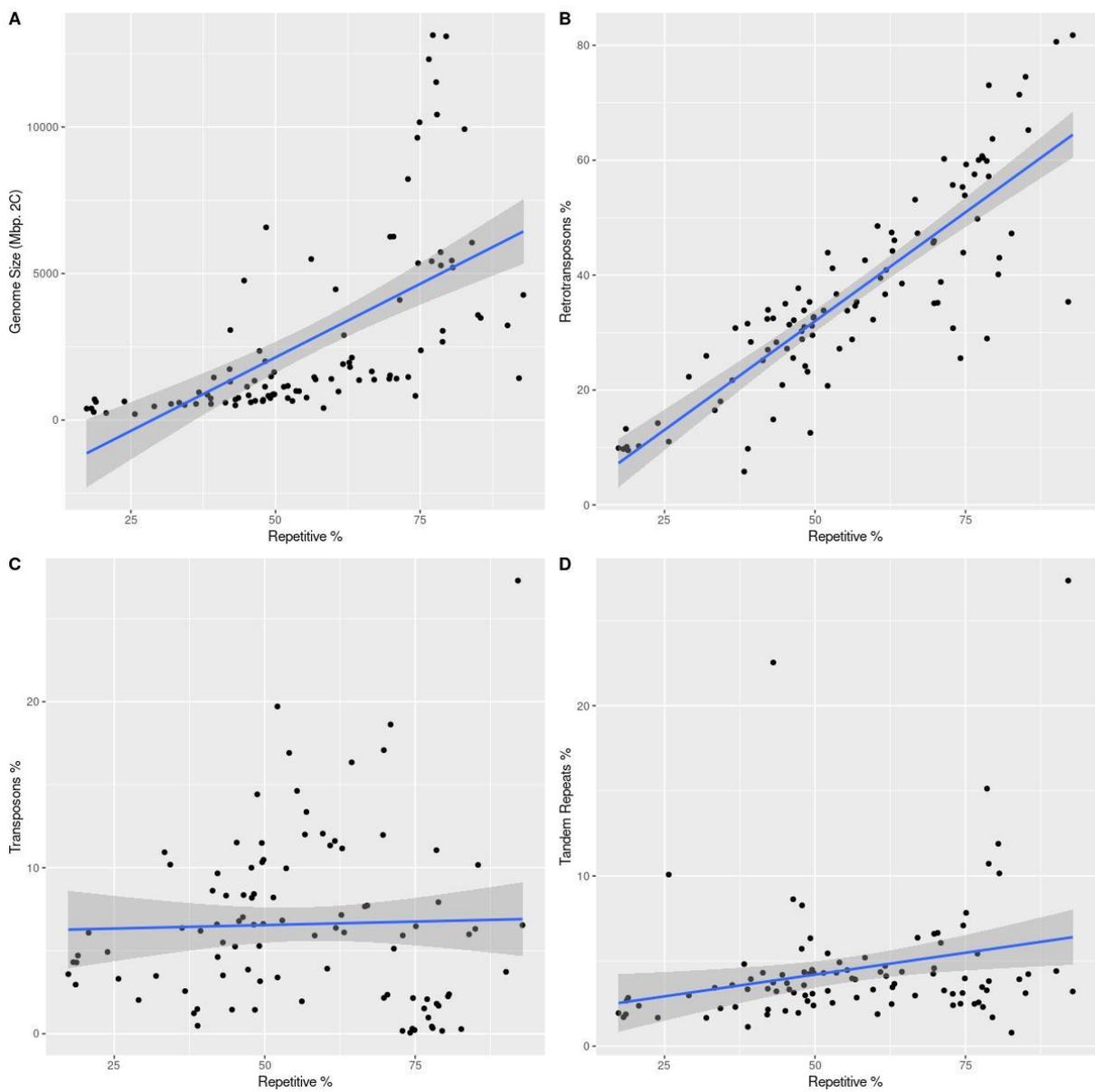


Figure 3. Different factors as function of repetitive frequency. (A) Genome size; (B) Retrotransposons; (C) Transposons and (D) Tandem Repeats.

Table 1. Summary of repetitive DNA sequences in *Syagrus coronata*.

Class	Order	Superfamily	Family	Genome [%]		
				<i>Syagrus coronata</i> (Ouricuri)		
Classified TEs Retrotransposon	LTR	Ty1/Copia	<i>Angela</i>	26.28		
			<i>Sire</i>	8.21		
			<i>Tork</i>	2.95		
			<i>Ikeros</i>	0.18		
					Total	37.62
		Ty3/Gypsy	<i>Athila</i>	0.11		
			<i>Ogre/Tat</i>	9.27		
			<i>CRM</i>	0.43		
			<i>Tekay</i>	2.41		
					Total	12.22
Classified TEs Transposon	DNA		<i>MuDR Mutator</i>	0.04		
			<i>Em/Spm/CACTA</i>	3.53		
			<i>hAT</i>	0.35		
			Total	3.92		
rDNA				0.56		
Repeats Classified				1.32		
Unclassified				4.75		
Total				60.39		

Table 2. Characterization of satellite sequences in *Syagrus coronata*.

Type		Size [pb]	CG frequency [%]	Repeat frequency [%]
DNA satellites	SatSCO2	110	64,5	0.260
	SatSCO4	184	69,6	0.094
	SatSCO5	276	28,3	0.093
	SatSCO6	163	26,4	0.073
	SatSCO11	193	23,0	0.017
Minisatellites	SatSCO3	60	41,7	0.150
	SatSCO1	55	56,4	0.760
	SatSCO7	93	49,5	0.071
	SatSCO8	26	57,7	0.021
	SatSCO9	21	33,3	0.020
	SatSCO10	56	67,9	0.020
Total				1.319

Table S1. Characterization of monomers of satellite sequences in *Syagrus coronata*.

Satellites	Monomers
SatSCO1	GGAGTCCGAATGGTTGAAGGAGTCCGGAGGCACTCTGACACTGTAGAA GTCCGGA
SatSCO2	AAGTCCGGCGGAGTCCGAATACCTGGAGGAGTCCGGCGGTGCTCCAGC ACTGTAGAAGTCCGGCGGAGTCCGAATGCCTGGAGGAGTCCGGCGGTG CTCCAGCACTGTAG
SatSCO3	CTGGAACAGAAACGTCCCCGAACGCAGAATTATGACATTAAGGGTCAA AATTATGACATC
SatSCO4	CTTGCTGTCGCGGGCTTGCCCGCAGGCCTGTCGCAGGGGGTGCCTGCGT GGGCCTGGCACGGTGTGCTGCGAAGGGGTTTCGGCTTCGTTGGGTTGCCAC GAAAATGCAATGCCGCTCGGTCATTTGGAACCCCCGGGGTGGGGCTTG CCCGGATGCCCCGCCGGGGTGGTAGGGGGTGTGCTGCTGC
SatSCO5	GAGCCAAAATTATGACATTCTGCATGTAAAACGTCTTTTGAAGGTAGTT TTGACAATTTTATAACATAGTAGATCACAACCATGACATTATATGTTAA TATTATGACATCTTGTAGGTAAAAGACCTTCCAAATATATTTTTTGATAT TTTTATAGCATTCTATGTCAAAAATATGGCATCCTGTAGGTAAAACAC CTCTCAAATATAGCTTCCAGCATTTTTATGACATCTAATGCCAAAATTA TGACATTTAAAGTCAAATTTATGACATTTA
SatSCO6	TCGAAATCCTGTCAAACCTAATTCAATTAGACTTTGAATCCAAGTCCTA CTTGGATTATAAAATCCAATTACCTTCAAATTAAGTCAAGCCTAATTA ATTTAATCTGATTTAAATCAATTAGAAGAAAATAATTTTTAATTTGGGT TCAATTCGAATCCAAT
SatSCO7	CTGAGTCACAATGAAGGCTGATTGAAGGTCTTCGGAAGGCTGTGGCCG AGGTCCTATTTAGGCCGAACCTTGAAGTTATCAAGAGATGAAGGC
SatSCO8	CCTCTGATCGGCCTAAGCTTCAACGG
SatSCO9	ATTTAAGACTGCACACAGTTA
SatSCO10	GTCCGGCCGAGAGTGCTCCTGCACTGTTGACATCCGGCCGGACGAGAG CCAGAGGG
SatSCO11	TTTTTTTTAAGGTTGATTTGTGTTTATTTTAGTAAACTTAGCATTATTTT TTATATTCATCTGGAATTATGGATCCAAATTTTAAATTTATACCATGTAG TTTCTGTTTATTTCAACAGAGTTTGTGTTTATCCTGCACAGTTTGTGTTT ATTGTTGCTATTTTATGTTTATTCGATGTTATTATTTAGTTTTTA

Table S2. Species and families used to correlate different factors as a function of repetitive frequency.

Family	Species	Mbp/2C	Repetitive	Retrotransposons	Transposons	Tandem repeat	Reference
Arecaceae	<i>Syagrus coronate</i>	4460	60,39	48,54	3,92	1,88	Presente estudo
Asperaceae	<i>Anacyclus clavatus</i>	10425,48	77,91	60,41	0,34	2,30	Vitales <i>et al.</i> (2020)
Asperaceae	<i>Anacyclus homogamos</i>	10161,42	74,90	53,87	0,22	3,98	Vitales <i>et al.</i> (2020)
Asperaceae	<i>Anacyclus linearilobus</i>	13134,54	77,19	60,03	0,97	2,57	Vitales <i>et al.</i> (2020)
Asperaceae	<i>Anacyclus maroccanus</i>	9633,3	74,52	55,33	0,30	3,12	Vitales <i>et al.</i> (2020)
Asperaceae	<i>Anacyclus monanthos</i>	11530,62	77,77	60,71	0,43	3,47	Vitales <i>et al.</i> (2020)
Asperaceae	<i>Anacyclus pyrethrum</i>	13095,42	79,49	63,71	0,17	1,69	Vitales <i>et al.</i> (2020)
Asperaceae	<i>Anacyclus radiatus</i>	12313,02	76,50	57,54	1,52	2,48	Vitales <i>et al.</i> (2020)
Asperaceae	<i>Anacyclus valentinus</i>	8224,98	72,89	55,69	0,17	3,07	Vitales <i>et al.</i> (2020)
Asperaceae	<i>Heliocauta atlantica</i>	9926,7	82,66	47,25	0,28	0,79	Vitales <i>et al.</i> (2020)
Paoceae	<i>Festuca pratensis</i>	6260	70,39	35,18	2,35	6,65	Zwyrtková <i>et al.</i> (2020)
Paoceae	<i>Festuca pratensis</i>	6254	69,81	35,09	2,16	6,60	Zwyrtková <i>et al.</i> (2020)
Paoceae	<i>Festuca arundinacea</i> ssp. <i>Arundinacea</i>	5494	56,17	28,8	1,95	3,98	Zwyrtková <i>et al.</i> (2020)
Paoceae	<i>Festuca arundinacea</i> ssp. <i>Glaucescens</i>	5276	78,58	28,95	1,81	15,13	Zwyrtková <i>et al.</i> (2020)

Paoceae	<i>Festuca gigantea</i>	6576	48,37	24,16	1,44	2,98	Zwyrtková <i>et al.</i> (2020)
Paoceae	<i>Festuca mairei</i>	4758	44,57	20,87	1,45	4,19	Zwyrtková <i>et al.</i> (2020)
Paoceae	<i>Lolium multiflorum</i>	5202	80,63	43,02	2,38	10,15	Zwyrtková <i>et al.</i> (2020)
Paoceae	<i>Lolium multiflorum</i>	5442	80,46	40,12	2,25	11,89	Zwyrtková <i>et al.</i> (2020)
Paoceae	<i>Lolium perenne</i>	5418	76,99	49,8	2,08	5,43	Zwyrtková <i>et al.</i> (2020)
Paoceae	<i>Lolium perenne</i>	5350	74,64	43,91	2,15	7,09	Zwyrtková <i>et al.</i> (2020)
Apocynaceae	<i>Hancornia speciosa</i>	860	38,24	5,7983	1,23	4,82	Santos <i>et al.</i> (2020)
Apocynaceae	<i>Asclepias syriaca</i>	822	74,18	25,535	0,063	2,49	Santos <i>et al.</i> (2020)
Apocynaceae	<i>Catharanthus roseus</i>	1486	49,23	12,5499	3,16	6,34	Santos <i>et al.</i> (2020)
Apocynaceae	<i>Rhazya stricta</i>	548	38,85	9,773	0,481	1,13	Santos <i>et al.</i> (2020)
Actinidiaceae	<i>Actinidia chinensis</i>	1307,86	42,163	27,0128	9,6597	3,37	Wang <i>et al.</i> (2020)
Amborellaceae	<i>Amborella trichopoda</i>	1412,66	70,9118	38,8032	18,623	6,07	Wang <i>et al.</i> (2020)
Bromeliaceae	<i>Ananas comosus</i>	763,8	55,3662	33,8178	14,6244	4,47	Wang <i>et al.</i> (2020)
Ranunculaceae	<i>Aquilegia coerulea</i>	603,96	45,7285	31,3773	6,7821	3,33	Wang <i>et al.</i> (2020)
Brassicaceae	<i>Arabidopsis thaliana</i>	239,34	20,7337	10,2516	6,0815	2,37	Wang <i>et al.</i> (2020)
Sapotaceae	<i>Argania spinosa</i>	1340,2	46,3699	25,5694	7,0203	8,63	Wang <i>et al.</i> (2020)
Asparagaceae	<i>Asparagus officinalis</i>	2375,08	75,118	59,262	6,4708	7,83	Wang <i>et al.</i> (2020)
Begoniaceae	<i>Begonia fuchsioides</i>	747,82	49,0972	35,3316	5,2858	4,25	Wang <i>et al.</i> (2020)

Amaranthaceae	<i>Beta vulgaris</i>	1132,36	51,4349	33,8493	8,2038	4,3	Wang <i>et al.</i> (2020)
Betulaceae	<i>Betula pendula</i>	871,82	49,5818	29,5396	10,33	3,08	Wang <i>et al.</i> (2020)
Gesneriaceae	<i>Boea hygrometrica</i>	3042,72	78,8918	73,0459	1,6838	3,82	Wang <i>et al.</i> (2020)
Urticaceae	<i>Boehmeria nivea</i>	689,24	47,853	28,8406	8,186	8,27	Wang <i>et al.</i> (2020)
Theaceae	<i>Camellia sinensis Kuntze</i>	5726,5	78,5357	59,865	11,0568	3,28	Wang <i>et al.</i> (2020)
Caricaceae	<i>Carica papaya</i>	739,56	38,7987	31,5508	1,4853	3,34	Wang <i>et al.</i> (2020)
Cactaceae	<i>Carnegiea gigantea</i>	1960,7	62,7399	47,4137	7,1533	2,47	Wang <i>et al.</i> (2020)
Cephalotaceae	<i>Cephalotus follicularis</i>	3229,04	90,0981	80,6082	3,7273	4,41	Wang <i>et al.</i> (2020)
Ceratophyllaceae	<i>Ceratophyllum demersum</i>	1466,52	72,9405	30,7579	5,919	2,4	Wang <i>et al.</i> (2020)
Amaranthaceae	<i>Chenopodium quinoa</i>	2667,1	78,8616	57,1767	7,9221	10,72	Wang <i>et al.</i> (2020)
Lauraceae	<i>Cinnamomum kanehirae</i>	1460,84	56,6838	34,6432	11,9992	3,92	Wang <i>et al.</i> (2020)
Rutaceae	<i>Citrus sinensis Osbeck</i>	655,34	46,4896	32,1453	8,3512	3,14	Wang <i>et al.</i> (2020)
Rubiaceae	<i>Coffea canephora</i>	1137,22	45,0774	35,0141	5,2523	2,07	Wang <i>et al.</i> (2020)
Cucurbitaceae	<i>Cucumis sativus</i>	387,66	17,3646	9,887	3,5889	1,94	Wang <i>et al.</i> (2020)
Datisceae	<i>Datisca glomerata</i>	1376,8	67,0502	47,2669	7,7208	6,37	Wang <i>et al.</i> (2020)
Apiaceae	<i>Daucus carota</i>	843	45,3483	27,1991	11,5161	3,7	Wang <i>et al.</i> (2020)
Caryophyllaceae	<i>Dianthus caryophyllus</i>	1135,32	48,1817	33,8572	6,5684	3,58	Wang <i>et al.</i> (2020)
Sapindaceae	<i>Dimocarpus longan</i>	990,66	53,5606	36,6937	9,962	4,31	Wang <i>et al.</i> (2020)

Arecaceae	<i>Elaeis guineensis</i>	3070,04	42,1908	33,9532	4,6224	2,15	Wang <i>et al.</i> (2020)
Myrtaceae	<i>Eucalyptus grandis</i>	1382,54	56,9231	35,3153	13,3568	2,85	Wang <i>et al.</i> (2020)
Nymphaeaceae	<i>Euryale ferox</i>	1450,46	39,3402	28,3573	6,1958	3,94	Wang <i>et al.</i> (2020)
Polygonaceae	<i>Fagopyrum esculentum</i>	2355,38	47,2315	37,7132	3,8606	1,95	Wang <i>et al.</i> (2020)
Oleaceae	<i>Fraxinus excelsior</i>	1734,92	42,0652	32,3716	6,5881	1,85	Wang <i>et al.</i> (2020)
Asteraceae	<i>Helianthus annuus</i>	6055,68	83,9536	71,4129	5,9867	3,93	Wang <i>et al.</i> (2020)
Cannabaceae	<i>Humulus lupulus</i>	4098,42	71,4615	60,2295	5,1251	3,27	Wang <i>et al.</i> (2020)
Convolvulaceae	<i>Ipomoea trifida</i>	984,76	54,067	27,1879	16,9133	4,92	Wang <i>et al.</i> (2020)
Juglandaceae	<i>Juglans regia</i>	1401,24	59,6524	32,2471	12,0523	3,32	Wang <i>et al.</i> (2020)
Crassulaceae	<i>Kalanchoe fedtschenkoi</i>	512,7	34,3001	18,0138	10,1881	2,22	Wang <i>et al.</i> (2020)
Cyperaceae	<i>Kobresia littledalei</i>	747,7	52,1066	20,7299	19,7091	5,45	Wang <i>et al.</i> (2020)
Linaceae	<i>Linum usitatissimum</i>	632,34	23,882	14,2324	4,9268	1,67	Wang <i>et al.</i> (2020)
Magnoliaceae	<i>Liriodendron chinense</i>	3484,84	85,4492	65,2387	10,1678	4,23	Wang <i>et al.</i> (2020)
Caprifoliaceae	<i>Lonicera japonica</i>	1807,62	62,8754	44,1966	11,1585	3,46	Wang <i>et al.</i> (2020)
Papaveraceae	<i>Macleaya cordata</i>	755,66	43,5867	28,3056	8,32	3,22	Wang <i>et al.</i> (2020)
Rosaceae	<i>Malus domestica</i>	1405,92	69,6597	45,6309	11,9717	4,25	Wang <i>et al.</i> (2020)
Euphorbiaceae	<i>Manihot esculenta</i>	1164,24	52,1242	43,9018	3,3894	3,25	Wang <i>et al.</i> (2020)
Fabaceae	<i>Medicago truncatula</i>	825,6	48,7616	23,1864	14,4172	2,65	Wang <i>et al.</i> (2020)

Lamiaceae	<i>Mentha longifolia</i>	706,58	18,7214	10,1134	4,2938	2,68	Wang <i>et al.</i> (2020)
Moringaceae	<i>Moringa oleifera</i>	2894	61,8276	40,8954	6,3722	4,12	Wang <i>et al.</i> (2020)
Moraceae	<i>Morus notabilis</i>	640,76	47,7926	30,2481	9,9985	5,72	Wang <i>et al.</i> (2020)
Musaceae	<i>Musa acuminata</i>	944,46	36,7722	30,7795	2,5663	2,3	Wang <i>et al.</i> (2020)
Nelumbonaceae	<i>Nelumbo nucifera</i>	1634,54	49,7532	32,5772	6,6142	2,39	Wang <i>et al.</i> (2020)
Nyssaceae	<i>Nyssa sinensis</i>	2002,9	48,2123	30,9752	8,414	4,35	Wang <i>et al.</i> (2020)
Araliaceae	<i>Panax notoginseng</i>	3581,52	84,9907	74,5124	6,3163	3,11	Wang <i>et al.</i> (2020)
Orchidaceae	<i>Phalaenopsis equestris</i>	2128,1	63,197	46,0523	6,1011	3,66	Wang <i>et al.</i> (2020)
Piperaceae	<i>Piper nigrum</i>	1522,44	69,7947	45,9508	17,0789	4,58	Wang <i>et al.</i> (2020)
Campanulaceae	<i>Platycodon grandiflorus</i>	1360,36	64,4508	38,5254	16,3444	4,37	Wang <i>et al.</i> (2020)
Salicaceae	<i>Populus trichocarpa</i>	868,26	49,5155	31,1591	11,4886	4,49	Wang <i>et al.</i> (2020)
Primulaceae	<i>Primula veris</i>	619,38	18,9583	9,4923	4,711	2,84	Wang <i>et al.</i> (2020)
Lythraceae	<i>Punica granatum</i>	592,76	41,3395	25,1467	8,6153	4,31	Wang <i>et al.</i> (2020)
Fagaceae	<i>Quercus suber</i>	1906,6	61,6596	36,6633	11,6114	4,71	Wang <i>et al.</i> (2020)
Apocynaceae	<i>Rhazya stricta</i>	548,7	31,9477	25,9334	3,4799	1,66	Wang <i>et al.</i> (2020)
Rhizophoraceae	<i>Rhizophora apiculata</i>	464,12	29,0432	22,2991	2,0295	2,98	Wang <i>et al.</i> (2020)
Santalaceae	<i>Santalum album</i>	392,2	18,188	9,731	4,31	1,7	Wang <i>et al.</i> (2020)
Pedaliaceae	<i>Sesamum indicum</i>	549,82	36,2644	21,7045	6,367	3,59	Wang <i>et al.</i> (2020)

Solanaceae	<i>Solanum lycopersicum</i>	1655,5	66,6373	53,1309	7,6647	2,97	Wang <i>et al.</i> (2020)
Araceae	<i>Spirodela polyrhiza</i>	273,34	18,5793	13,2391	2,958	1,87	Wang <i>et al.</i> (2020)
Cleomaceae	<i>Tarenaya hassleriana</i>	499,86	43,0548	32,475	5,4973	3,73	Wang <i>et al.</i> (2020)
Malvaceae	<i>Theobroma cacao</i>	649,44	52,9128	41,1864	6,8234	2,54	Wang <i>et al.</i> (2020)
Dioscoreaceae	<i>Trichopus zeylanicus</i>	1426,82	110,0194	35,3483	27,2885	27,35	Wang <i>et al.</i> (2020)
Celastraceae	<i>Tripterygium wilfordii</i>	697,06	43,067	14,8779	3,5149	22,54	Wang <i>et al.</i> (2020)
Lentibulariaceae	<i>Utricularia gibba</i>	201,38	25,696	11,0205	3,3086	10,08	Wang <i>et al.</i> (2020)
Vitaceae	<i>Vitis vinifera</i>	970,66	60,8775	39,4987	11,3409	4,36	Wang <i>et al.</i> (2020)
Velloziaceae	<i>Xerophyta viscosa</i>	590,92	33,357	16,482	10,9287	3,43	Wang <i>et al.</i> (2020)
Rhamnaceae	<i>Ziziphus jujuba</i>	875,5	49,8045	32,736	10,4757	4,29	Wang <i>et al.</i> (2020)
Zosteraceae	<i>Zostera marina</i>	407,82	58,3071	42,5747	5,9149	5,2	Wang <i>et al.</i> (2020)
Poaceae	<i>Zea mays</i>	4268,74	92,8326	81,757	6,5389	3,21	Wang <i>et al.</i> (2020)

CAPÍTULO III

Diversidade genética em *Syagrus coronata* (Mart.) Becc. (Arecaceae): implicações para manejo e conservação genética

Manuscrito a ser enviado para publicação na Revista **Crop Breeding and Applied Biotechnology**.

Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.

Diversidade genética em *Syagrus coronata* (Mart.) Becc. (Arecaceae): implicações para manejo e conservação genética

Jackeline Terto da Silva Santana^{1,2}, Marislane Carvalho Paz de Souza¹, Gleica Maria Correia Martins¹, Cícero Almeida¹ e Reginaldo de Carvalho²

¹*Laboratório de Recursos Genéticos, Universidade Federal de Alagoas, Campus de Arapiraca, Brasil*

²*Laboratório de Citogenética, Universidade Federal Rural de Pernambuco, Brasil*

Autor para correspondência:

Cícero Almeida

Campus de Arapiraca

Universidade Federal de Alagoas

Avenida Manoel Severino Barbosa s/n Rodovia AL 115, km 6,5, Bairro Bom Sucesso

Arapiraca, AL, Brasil

Telefone: (82) 3482-1831

E-mail: cicero.arapiraca@ufal.br

ORCID: <https://orcid.org/0000-0002-6793-8787>

Resumo

Syagrus coronata (Mart.) Becc. é uma palmeira endêmica do Brasil, distribuída nos biomas da Caatinga e Mata Atlântica com elevada importância socioeconômica e ecológica. No entanto, a acentuada fragmentação dos biomas onde é encontrada tem contribuído para a vulnerabilidade da espécie que necessita de estratégias para conservação de seus recursos genéticos. Diante disso, objetivou-se estimar o polimorfismo em indivíduos provenientes de seis localidades em diferentes biomas com base em marcadores microssatélites e sequências de cloroplasto. As análises revelaram a formação de apenas um grupo populacional, indicando baixa diferenciação genética entre as localidades. A maior parte da variação genética encontra-se dentro das amostras (98,3% para dados de SSRs e 56,68% para SNPs), revelando elevado fluxo gênico intrapopulacional. Os resultados encontrados expressam níveis significativos de diversidade genética podendo contribuir para o desenvolvimento de estratégias de conservação da espécie.

Palavras-chave: palmeira nativa, ouricuri, marcadores moleculares, conservação.

Introdução

A conservação dos recursos genéticos tem sido uma necessidade para muitas espécies, devido a perda de hábitat, superexploração dos recursos naturais, invasão de espécies exóticas, poluição, mudanças climáticas e outras ações antrópicas, que aumentam o risco de extinção da espécie, sendo a quantificação da diversidade genética e o grau de endogamia em populações, importantes para caracterizar a história da seleção natural e os relacionamentos entre os indivíduos da espécie. Esses parâmetros genéticos são fundamentais para o planejamento da conservação dos recursos genéticos (Hohenlohe et al. 2020). As estimativas de parâmetros de diversidade genética têm aumentado com a descoberta e o desenvolvimento de marcadores moleculares. Os marcadores moleculares têm sido acessíveis a muitas espécies depois do advento das tecnologias de sequenciamento de nova geração (NGS), permitindo disponibilizar diversos tipos de marcadores moleculares, como os marcadores microsatélites (SSR – *Simple Sequence Repeat*) (Hedel et al. 2016, Moura et al. 2017) e mais recentemente os SNPs (*Single Nucleotide Polymorphism*) (Morgil et al. 2020), que possui abundância nos genomas, facilidade nas análises e gerenciamento de dados (Thomson 2014).

Syagrus coronata (Mart.) Becc. é uma espécie endêmica do Brasil pertencente a Arecaceae, sendo popularmente conhecida como ouricuri e licuri (Noblick 2017). Trata-se de uma palmeira com elevado potencial socioeconômico e ecológico devido à vasta utilização para fabricação de produtos alimentícios e artesanais (Noblick 2017), além de seus frutos servirem como alimento para a fauna nativa, principalmente a ave *Anodorhynchus leari* que se encontra ameaçada de extinção (Santos-Neto e Camandaroba 2008). Essa espécie está distribuída em domínios da Mata Atlântica e Caatinga (Noblick 2017, Souza et al. 2018), considerados biomas de grande importância para o Brasil, com elevado grau de endemismo. No entanto, ambos vêm sofrendo acentuada fragmentação de hábitat ocasionando a redução de sua cobertura vegetal. O primeiro se encontra com apenas 11% a 16% de sua extensão original (Ribeiro et al. 2009) e o segundo possui menos de 10% de sua área de origem, reconhecido como uma das regiões naturais mais ameaçadas do país (Banda et al. 2016, Moro et al. 2016).

Estudos genéticos vêm sendo realizadas em *S. coronata*, como o sequenciamento do genoma de cloroplasto (Áquila et al. 2018), desenvolvimento de marcadores moleculares do tipo SSR (Barbosa e Almeida 2018) e a estrutura filogeográfica nos domínios da Mata atlântica e Caatinga (Souza et al. 2018). Esses estudos filogeográficos revelaram expressiva variabilidade genética nos haplótipos, sugerindo elevada diversidade genética nas populações naturais que pode ser utilizada para elaboração de estratégias para formação de bancos de germoplasma. No presente estudo, foi acessada a diversidade genética em *S. coronata* usando marcadores moleculares SSR e sequências do espaçador intergênico do genoma de cloroplasto *trnH-psbA*, com os seguintes objetivos: (1) obter o polimorfismo de *loci* de microssatélites e sequências do genoma de cloroplasto em *S. coronata* e (2) propor estratégias para coleta e conservação da espécie.

Material e métodos

Material vegetal e extração do DNA

A coleta do material vegetal consistiu em amostras foliares em estágio adulto e sem sinais de injúria pertencentes a seis localidades, sendo quatro da Caatinga (São José da Tapera, Santana do Ipanema, Raso da Catarina, Água Branca) e duas da Mata Atlântica (Pirambú e Coruripe), onde o número de indivíduos amostrados variou entre 13 a 20, devidamente georreferenciados (Tabela S1). O isolamento do material genético ocorreu de acordo com o protocolo Doyle e Doyle (1987) e a integridade e quantidade do DNA foi aferida em gel de agarose a 1% e espectrofotômetro, respectivamente.

PCR e genotipagem das sequências de microssatélites (SSRs)

A amplificação do DNA ocorreu para cinco *loci* do genoma nuclear (SYA92, SYA100, SYA116, SYA129, SYA130), usando *primers* desenvolvidos por Barbosa e Almeida (2018). As reações de PCR foram conduzidas em um volume de 50 µL, para cada reação foi utilizado 100-200 ng de DNA, 5 µL da solução tampão (Tris-Hcl pH 8,5 e 500 mM de KCl), 0,2 mM de cada dNTP, 1,5 mM de MgCl₂, 0,5 µL de cada primer e 2,5 unidades de Taq Polimerase. As amplificações ocorreram em um termocilador (Applied Thermal Cycler Biosystems® 2720) com uma desnaturação inicial de 94° C por dois minutos, seguido por 35

ciclos de desnaturação a 94° C por 45 segundos, anelamento por 30 segundos de acordo com a temperatura específica para cada primer e extensão de 72° C por um minuto. Uma extensão final de 72° C por 10 minutos foi realizada. Os produtos de PCR foram analisados em gel de agarose a 1%. Para a genotipagem foi utilizado por amostra 1 µL da reação de PCR, 0,2 µL do marcador LIZ600 e 8,8 µL de formamida. Os produtos dessa reação foram submetidos à desnaturação a 95° C por dois minutos em termociclador e genotipados em sequenciador automático ABI3500 (Applied Biosystems).

PCR e sequenciamento das sequências de cpDNA

Para as regiões do genoma de cloroplasto foi realizada a amplificação de oito regiões (*rpL2*, *trnH-psbA*, *rps16-trnQ*, *trnD-trnT*, *rbclA1*, *trnS-trnG*, *petB-petD* e *atpB-rbcL*) usando *primers* descritos por Scarcelli et al. (2011), porém apenas a região *trnH-psbA* apresentou polimorfismo e as sequências foram usadas para genotipagem por sequenciamento de Sanger. As reações ocorreram com um volume de 50 µL, contendo 100-200 ng de DNA, 5 µL da solução tampão, 1,5 mM de MgCl₂, 0,2 mM dNTP, 1,25 U Taq DNA polymerase e 0,5 µL de cada *primer*. As amplificações foram realizadas com uma desnaturação inicial de 94° C durante três minutos, seguidas de 40 ciclos de 94° C por 30 segundos, anelamento de 55-60° C por 30 segundos e uma extensão final de 72° C por 10 minutos. As reações foram conduzidas em um termociclador e os produtos de PCR foram submetidos à eletroforese em gel de agarose a 1%. Os produtos de PCR foram sequenciados usando o kit de sequenciamento BigDye® Terminator v3.1 (Applied Biosystems®) no analisador genético ABI3500 (Applied Biosystems®).

Análise dos dados

Para os dados do *trnH-psbA* foram calculados os parâmetros de diversidade genética: índice de diversidade de Shannon-Wiener (H) (2001), índice de Stoddart e Taylor's (G) (1987), índice de diversidade de Simpsons (lambda) (Simpson 1949), riqueza de alelos, medida de informação de Shannon e heterozigosidade (Sherwin et al. 2017). Essas estatísticas foram calculadas usando o pacote poppr R package (Kamvar et al. 2014). Para os dados de SSR, foram analisados o número de alelos (Na) heterozigosidade observada (Ho) e esperada (He), F_{ST}, diversidade de Nei (1978) (H_{exp}) e a frequência de alelos nulos para cada locus foi

calculada usando Chakraborty e Zhong (1994), implementado no R PopGenReport 3.0 package (Gruber e Adamack 2017). Foram calculados os parâmetros de diversidade genética: índice de diversidade de Shannon-Wiener (H), índice de Stoddart e Taylor's (G), índice de diversidade de Simpsons (lambda) (Simpson 1949), riqueza de alelos e diversidade de Nei (1978) (Hexp) usando o pacote poppr R package (Kamvar et al. 2014). A Análise de componentes principais foi obtida usando a função glPca do pacote adegenet (Jombart 2008) e a análise de variância molecular usando poppr.amova do pacote poppr R package (Kamvar et al. 2014).

Resultados e discussão

Marcadores SNP

Entre as regiões do genoma de cloroplasto amplificadas, somente a região do *trnH-psbA* mostrou polimorfismo, exibindo cinco sítios polimórficos, combinados em nove haplótipos. Os haplótipos foram identificados em Algarismos Romanos variando de I a IX, cuja distribuição dos haplótipos V e VIII são os mais frequentes na Mata Atlântica, enquanto os haplótipos de maior frequência na Caatinga são os IX, V e VII. Embora o haplótipo VIII esteja presente apenas na população de Coruripe e o haplótipo VI sendo identificado apenas na população proveniente de São José da Tapera, houve um compartilhamento de haplótipos expressivo entre os diferentes biomas (I, III, IV, V, VII e IX), indicando a ocorrência de fluxo gênico entre indivíduos da Caatinga e Mata Atlântica (Figura 1A).

A avaliação da diversidade genética nos marcadores SNPs usando diferentes estatísticas (H, G e lambda) demonstra que a população de Pirambú (bioma Mata Atlântica) possui os maiores valores em relação às demais localidades, porém a população coletada no Raso da Catarina (bioma Caatinga) apresentou maior diversidade usando o método Shannon information, além de também representar a população com o maior índice de heterozigotidade. A maior riqueza alélica foi encontrada em São José da Tapera e Pirambú, ambas com 0,8 (Tabela 1). Vale a pena considerar que a presença de fluxo gênico pode ser por consequência da eficiência nos mecanismos de dispersão das sementes. Apesar de alguns estudos apontarem *A. leari* como agente de dispersão de sementes, esta ave aparentemente não desempenha este papel, visto que consome os frutos imaturos (Andrade et al. 2015). Dessa forma, a maneira mais provável de dispersão das sementes nos diferentes biomas foi

realizada por animais terrestres e pela atuação do homem, transportando as sementes de um local para o outro (Souza et al. 2018).

Marcadores SSR

Já para as regiões de microssatélites genotipadas, apenas três apresentaram polimorfismo (P92, P129 e P130). O número de alelos encontrados de acordo com cada localidade variou de 15 a 18 alelos, representando bem a diversidade nas localidades estudadas (Tabela 3). Apenas na Loc.D não foi identificada presença de alelo restrito, as demais apresentaram a seguinte exclusividade alélica: Loc.A e Loc.C apresentaram alelos restritos para o locus P130; Loc.B para o locus P92 e Loc.E e Loc.F para os loci P92 e P130 (Figura 1B, C e D e Tabela 4), sugerindo que estas localidades preservam seu próprio pool gênico.

A análise dos SSRs revelou que a maior diversidade genética foi encontrada na população de Água Branca (bioma Caatinga) para os três métodos estatísticos utilizados (H, G e lambda). Já a diversidade baseada em Nei (1978) indica que a população de Pirambú foi mais expressiva e também apresentou a maior diversidade alélica (5,09) (Tabela 4).

Os níveis de diversidade genética foram elevados para todos os *loci* analisados, onde a heterozigotidade observada (HO) obteve valores superiores a 0,67 e a heterozigotidade esperada (HE) valores acima de 0,58, de modo que $HO > HE$ pelo equilíbrio de Hardy-Weinberg, revelando excesso de heterozigotos. A heterozigotidade baseada em Nei (1978) variou de 0,59 a 0,78, também indicando variabilidade genética. O maior valor obtido para F_{ST} nos três *loci* foi 0,048, evidenciando diferenciação genética reduzida entre as localidades (valor distante de 1) e que apenas um grupo foi definido. O valor de alelos nulos variou de 0,09 a 0,19 para P129 e P130, respectivamente (Tabela 3), sendo que a Loc.F apresentou a maior taxa de falha de amplificação (PCR) (52,63%) no locus P129, indicando que determinada região deste locus para esta população apresenta uma alteração genética na região de anelamento que impossibilita o sucesso da reação em cadeia da polimerase (PCR) (Figura S1).

Diversidade Genética nas Populações

A análise de variância (AMOVA) demonstrou que a diversidade genética se encontra dentro das amostras e não entre as amostras tanto para dados oriundos de SSRs (98,3%) quanto para as sequências de cloroplasto (56,68%) (Tabelas 2 e 5), indicando um elevado fluxo gênico intrapopulacional. Embora os recursos genéticos da espécie estejam distribuídos em diferentes biomas (Souza et al. 2018), este resultado evidencia que a diversidade presente nas localidades estudadas não difere entre si, ou seja, para a formação de bancos de germoplasma não há a necessidade de coletar indivíduos em locais distintos. Essa evidência foi corroborada através das análises de componentes principais (PCA) e rede mínima de agrupamento em que detectaram a formação de um único grupo, indicando que a variação genética não difere entre as localidades estudadas (Figura 2). Apesar das localidades apresentarem expressiva aproximação genética, formando apenas um grupo, pode-se notar que todas as localidades com vegetação da Caatinga parecem estar mais próximas, com exceção da Loc.D (Figura S2).

Ao analisar pares de grupos populacionais, o índice de distância genética revelou baixa diferenciação genética entre as localidades para os dois tipos de marcadores, no qual o maior valor obtido para SSRs esteve entre Raso da Catarina e Santana do Ipanema (0,101) e o menor valor foi encontrado entre Água Branca e Raso da Catarina (0,023). Para os dados de SNPs, o maior valor foi entre Água Branca e Raso da Catarina (0,755) e o menor valor obtido foi entre Santana do Ipanema e Coruripe (-0,107) (Tabela 6).

Conservação de recursos genéticos

Considerando que uma das principais metas nos programas de conservação, adicionalmente a preservação de habitats, é a manutenção dos níveis existentes de diversidade genética em espécies ameaçadas ou que estão inseridas em biomas ameaçados, os resultados obtidos neste estudo possibilitaram a compreensão acerca da diversidade genética de *S. coronata* em seis diferentes localidades distribuídas pelos biomas Caatinga e Mata Atlântica contribuindo para a conservação dos recursos genéticos da espécie.

Partindo do princípio que a palmeira ouricuri apresenta monoicismo, possuindo flores unissexuadas (configurando uma distribuição padrão, onde as flores pistiladas estão sempre entre duas estaminadas) (Noblick 2017), sendo considerada xenógama facultativa, algumas características fenotípicas da planta podem não ser mantidas devido à possível recombinação genética. A formação de bancos de sementes pode ser viável, pois as sementes

de ouricuri são ortodoxas, apesar de estudos de germinação na espécie serem incipientes. No entanto, a conservação *in situ* parece ser a opção mais favorável, devido à alta diversidade genética encontrada dentro das populações, indicando que a diversidade se encontra amplamente distribuída nas diferentes localidades estudadas. Além disso, é sugerido promover o enriquecimento das Unidades de Conservação (UC) que contam com a presença da palmeira ouricuri como, por exemplo, parques (nacionais, estaduais e municipais), estação ecológica e áreas de proteção ambiental (Gomes et al. 2018).

De maneira geral, pode-se afirmar que o conhecimento acerca da importância da conservação de *S. coronata* merece ser melhor difundido na sociedade, merecendo maior aproximação entre comunidades rurais, instituições de pesquisa e o consumidor propriamente dito, a fim de atingir resultados cada vez mais promissores. Dessa forma, deve-se levar em consideração as práticas extrativistas locais, principalmente aquelas voltadas para a extração de folhas para confecção artesanal (Andrade et al. 2018), que é responsável pela sobrevivência de diversas comunidades (Noblick 2017). Neste contexto, o entendimento sobre as práticas de colheita de folhas pode auxiliar para a realização de medidas mais sustentáveis, podendo contribuir para melhores estratégias de conservação (Campos et al. 2018),

Agradecimentos

Nós agradecemos à Universidade Federal de Alagoas (UFAL), pelos laboratórios utilizados e pelo suporte científico, à Fundação de Amparo à pesquisa do Estado de Alagoas (FAPEAL) por financiar este projeto e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de doutorado.

Conflitos de interesse

Os autores declaram não haver conflito de interesse.

Referências

- Adamack AT and Gruber B (2017) PopGenReport: simplifying basic population genetic analyses in R. **Methods in Ecology and Evolution** 5:384-387.
- Andrade WM, Ramos MA, Souto WMS, Bento-Silva JS, Albuquerque UP and Araújo EL (2015) Knowledge, uses and practices of the licuri palm (*Syagrus coronata* (Mart.) Becc.) around protected areas in northeastern Brazil holding the endangered species Lear's Macaw (*Anodorhynchus leari*). **Tropical Conservation Science** 8:893-911.
- Áquila R, Silva J, Silva JV and Almeida C (2018) The first complete chloroplast genome sequence of *Syagrus coronata*: comparative analysis in Arecoideae (Arecaceae). **Brazilian Journal of Botany** 41:393-401.
- Banda K, Delgado-Salinas A, Dexter KG, Linares-Palomino R, Oliveira-Filho A, Prado D, Pullan M, Quintana C, Riina R, Rodríguez GM, et al. (2016) Plant diversity patterns in neotropical dry forests and their conservation implications. **Science** 353:1383-1387.
- Barbosa SN and Almeida C (2018) Genome survey and development of 15 microsatellite molecular markers in *Syagrus coronata* (Martius) Beccari (Arecaceae) by next-generation sequencing. **Brazilian Journal of Botany** 42:195-200.
- Campos JLA, Araújo EL, Gaoue OG and Albuquerque UP (2018) How can local representations of changes of the availability in natural resources assist in targeting conservation? **Science of the Total Environment** 628(629): 642-649.
- Chakraborty R and Zhong Y (1994) Statistical power of an exact test of Hardy-Weinberg proportions of genotypic data at a multiallelic locus. **Human Heredity** 44:1-9.
- Doyle JJ and Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Phytochemical Bulletin** 19:11-15.
- Gomes FM, Ribeiro RTM, Batista MEP, Souza EB, Siqueira Filho JA, Portela S, Castro R and Loiola MIB (2018) *Syagrus coronata* - Licuri. In: Coradin L, Camillo J and Pareyn FGC (eds) **Espécies nativas da flora brasileira de valor econômico atual ou potencial – plantas para o futuro – região Nordeste**. Ministério do Meio Ambiente, Brasília, p. 1056-1072.
- Hodel RGJ, Segovia-Salcedo C, Landis JB, Crowl AA, Sun M, Liu X, Gitzendanner MA, Douglas NA, Gernam-Aubrey CC, Chen S, Soltis DE, et al. (2016) The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. **Applications in Plant Sciences** 4:1600025.
- Hohenlohe PA, Funck WC and Rajora OP (2020) Population genomics for wildlife conservation and management. **Molecular Ecology** 30:62-82.
- Jombart T (2008) ADEGENET: a R package for the multivariate analysis of genetic markers. **Bioinformatics** 24:1403-1405.
- Kamvar ZN, Tabima JF and Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. **PeerJ** 2:e281.

Santana JTS (2022) Caracterização genômica do DNA repetitivo e diversidade genética em *Syagrus coronata* (Mart.) Becc.

Morgil H, Gercek YC and Tulum I (2020) Single nucleotide polymorphisms (SNPs) in plant genetics and breeding. In: Çalışkan M, Erol O e Öz GC (eds) **The recent topics in genetic polymorphisms**. IntechOpen, London, 12p.

Moro MF, Lughadha EN, Araújo FS and Martins FR (2016) A phylogeographical metaanalysis of the Semiarid Caatinga Domain in Brazil. **Botanical Review** **82**:91-148.

Moura RF, Dawson DA and Nogueira DM (2017) The use of microsatellite markers in neotropical studies of wild birds: a literature review. **Anais da Academia Brasileira de Ciências** **89**:145-154.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. **Genetics** **89**:583-590.

Noblick LR (2017) A revision the genus *Syagrus* (Arecaceae). **Phytotaxa** **294**:1-262.

Ribeiro MC, Metzger JP, Martensen AC, Ponzoni FJ and Hirota MM (2009) The Brazilian Atlantic Forest: How much is left, and now is the remaining forest distributed? Implication for conservation. **Biological Conservation** **142**:1141-1153.

Santos-Neto JR and Camandaroba M (2008) Mapeamentos dos sítios de alimentação da arara-azul-de-Lear (*Anodorhynchus leari*) (Bonaparte, 1856). **Ornithologia** **3**:1-17.

Scarcelli N, Barnaud A, Eiserhardt W, Treier UA, Seveno M, d'Anfray N, Vigouroux Y and Pintaud J-C (2011) A set of 100 chloroplast DNA primer pairs to study populations genetics and phylogeny in monocotyledons. **Plos ONE** **6**:e19954.

Shannon CE (2001) A mathematical theory of communication. **Mobile Computing and Communications review** **5**:3-55.

Sherwin WB, Chao A, Jost L and Smouse PE (2017) Information theory broadens the spectrum of molecular ecology and evolution. **Trends in Ecology & Evolution** **32**:1-16.

Simpson EH (1949) Measurement of diversity. **Nature** **163**:688.

Souza MCP, Moura F, Silva JV and Almeida C (2018) Phylogeography of the palm *Syagrus coronata* (Martius) Beccari (Arecaceae): distribution in the “Caatinga” and Atlantic Forest domains. **Brazilian Journal of Botany** **41**:849-857.

Stoddart JA and Taylor JF (1987) Genotypic diversity: estimation and prediction in samples. **Genetics** **118**:705-711.

Thomson MJ (2014) High-throughput SNP genotyping to accelerate crop improvement. **Plant Breeding and Biotechnology** **2**:195-212.

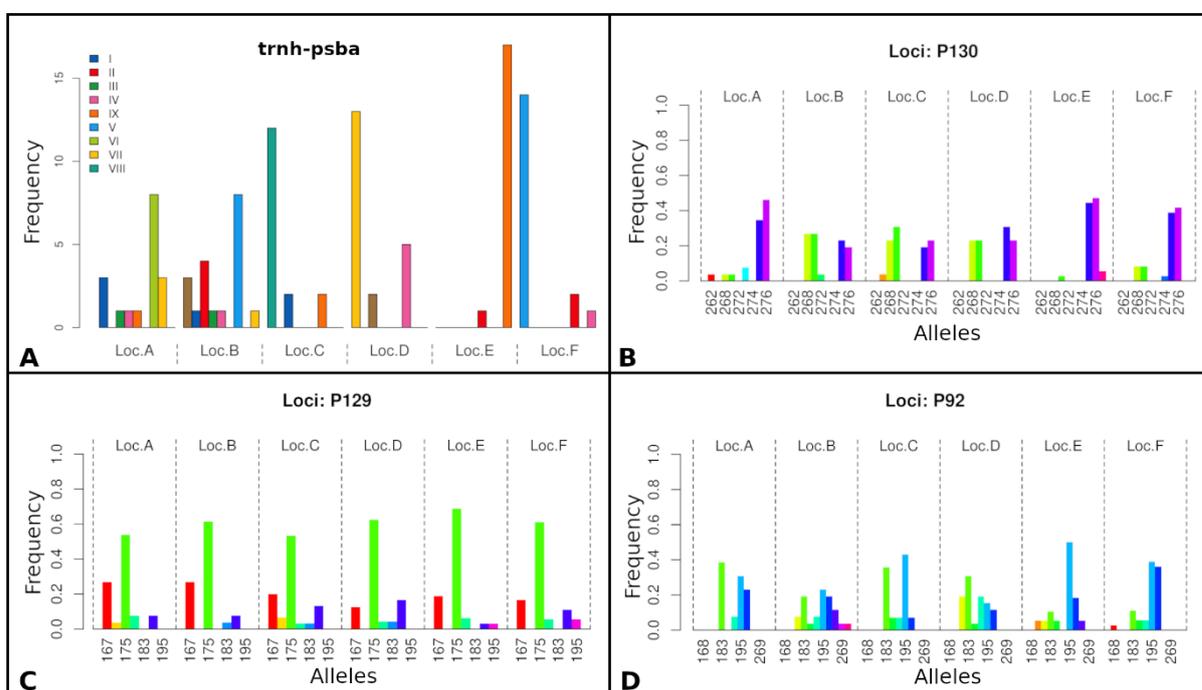


Figure 1. (A) Distribution of haplotypes from the *trnH-psbA*. The indicated colors indicate the different haplotypes found and e (B), (C) e (D) Allelic distribution in three microsatellite regions. The abbreviations specified refer to the localities: Loc. A (São José da Tapera), Loc. B (Pirambú), Loc. C (Coruripe), Loc. D (Santana do Ipanema), Loc. E (Raso da Catarina) and Loc. F (Água Branca).

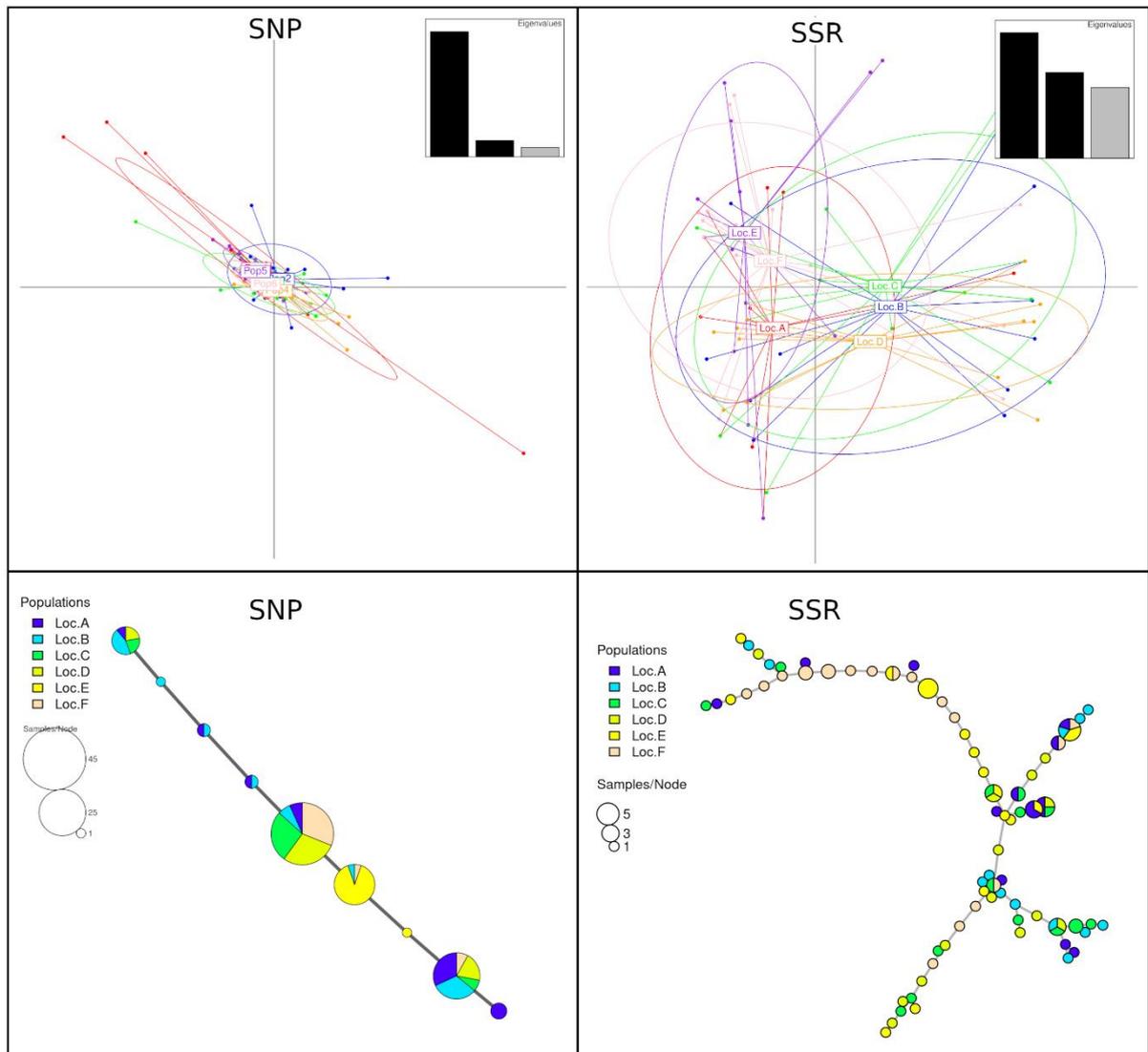


Figure 2. Principal component analysis (PCA) revealing a two-dimensional distribution using data of SNP and microsatellites. Minimum-spanning network depicting the distribution multilocus SNP and microsatellites in *Syagrus coronata*.

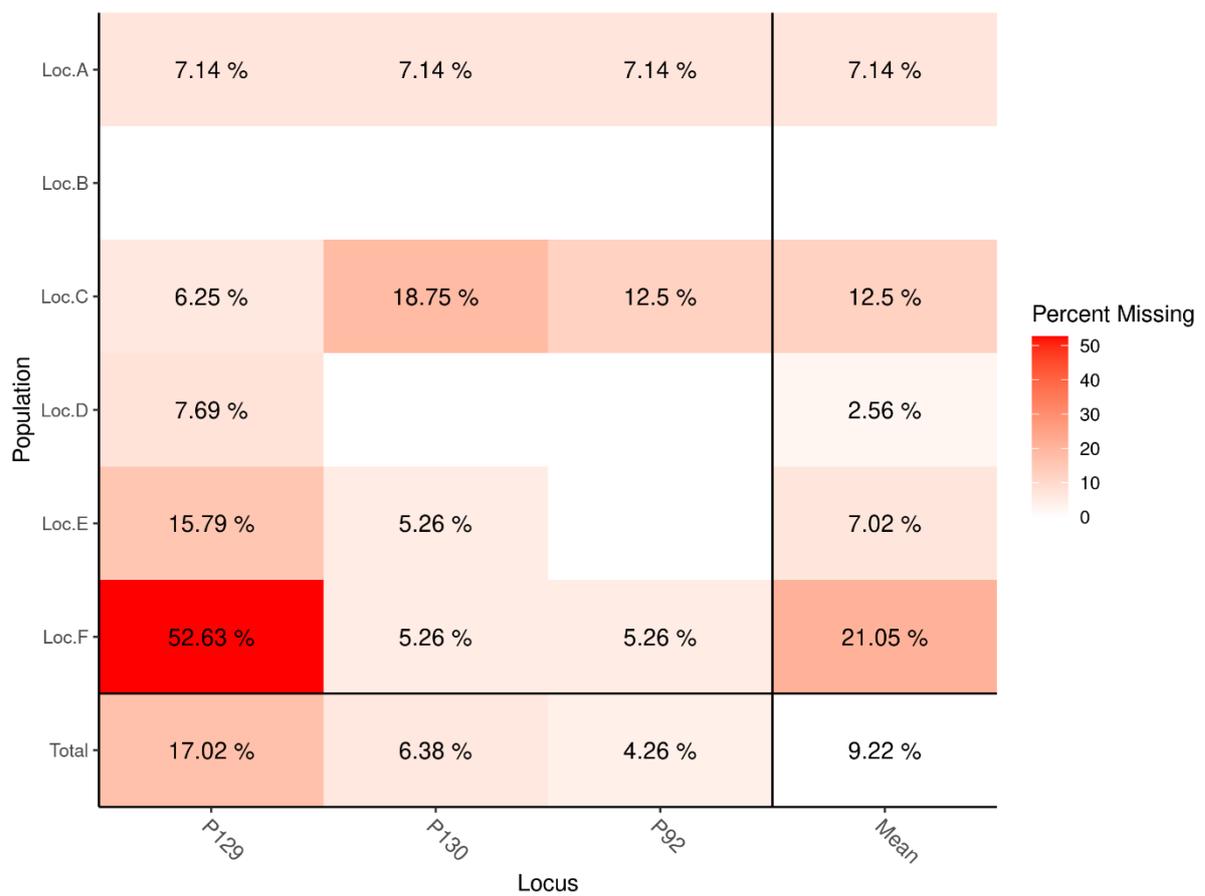


Figure S1. Percent missing during polymerase chain reaction (PCR) in diferent *loci*.

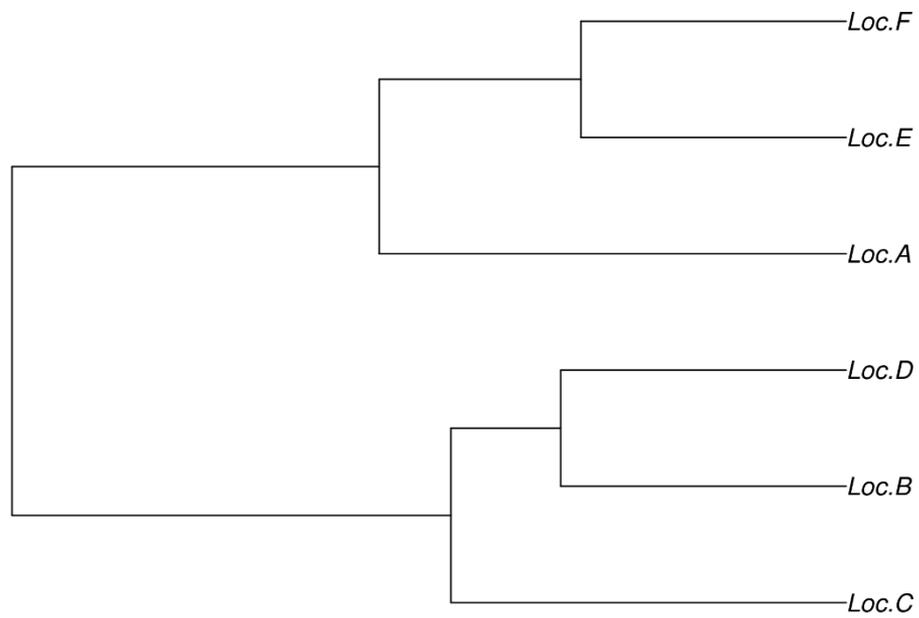


Figure S2. Clustering obtained from microsatellite sequences (SSRs) using the Neighbor Joining method.

Table 1. Genetic diversity parameters for *Syagrus coronata* using sequences data.

Populations	SNP loci						
	Samples	H	G	lambda	Allelic richness	Shannon information	Het
São José da Tapera (A)	17	1.47	3.40	0.71	0.8	0.203	0.082
Pirambú (B)	19	1.60	3.88	0.74	0.8	0.235	0.103
Coruripe (C)	16	0.73	1.68	0.41	0.4	0.179	0.203
Santana do Ipanema (D)	20	0.86	2.02	0.50	0.4	0.166	0.190
Raso da Catarina (E)	18	0.21	1.12	0.11	0.4	0.277	0.273
Água Branca (F)	17	0.58	1.44	0.30	0.4	0.164	0.216

H: Shannon-Wiener Index of MLG diversity (Shannon 2001); G: Stoddart and Taylor's Index of MLG diversity (Stoddart and Taylor 1988); Lambda: Simpson's Index (Simpson 1949), Allelic richness, Shannon information and het: heterozygosity base in Sherwin et al. (2017).

Table 2. Analysis of Molecular Variance between samples using sequences data from chloroplast.

Source of variation	D.F.	S. S.	V. C.	P. V.
Among samples*	5	24.47	0.26	43.32
Within samples	101	33.83	0.33	56.68
Total	106	58.30	0.59	100%

D.F.- Degree of freedom; S.S.- Sum of squares; V.C.- Components of Variance; P.V.- Percentage of variation

*Significant at $p < 0.001$ (999 permutation)

Table 3. Genetic diversity parameters for *Syagrus coronata*.

<i>Primers*</i>	Microsatellite loci							
	N. Alleles	Alleles size	Ta (°C)	H_O	H_E	Hexp	F_{ST}	Null alleles
P129	7	175–189	60	0.67	0.58	0.59	-0.012	0.09
P130	10	253–260	60	0.95	0.73	0.74	0.048	0.19
P92	11	189–212	60	0.80	0.77	0.78	0.037	0.10

* H_O : Observed heterozygosity; H_E : expected heterozygosity; Hexp: Nei's 1978 gene diversity.

Table 4. Genetic diversity parameters for *Syagrus coronata*.

Populations	Microsatellite loci							
	Samples	N° Alleles	Private alleles	Allelic richness	H	G	lambda	Hexp
São José da Tapera (A)	14	15	1	4.21	2.44	10.9	0.91	0.69
Pirambú (B)	13	18	2	5.09	2.56	13.0	0.91	0.74
Coruripe (C)	16	16	1	4.57	2.67	14.2	0.93	0.72
Santana do Ipanema (D)	13	15	0	4.54	2.63	13.0	0.92	0.73
Raso da Catarina (E)	19	16	2	4.13	2.57	10.9	0.91	0.60
Água Branca (F)	19	18	2	4.43	2.80	15.7	0.94	0.67

H: Shannon-Wiener Index of MLG diversity (Shannon 2001); G: Stoddart and Taylor's Index of MLG diversity (Stoddart and Taylor 1988); Lambda: Simpson's Index (Simpson 1949), Allelic richness, Hexp: gene diversity (Nei 1978).

Table 5. Analysis of Molecular Variance between samples using SSR data.

Source of variation	D.F.	S. S.	V. C.	P. V.
Among populations*	5	9.21	0.025	1.7%
Within populations	88	127.71	1.45	98.3%
Total	93	136.92	1.48	100%

D.F.- Degree of freedom; S.S.- Sum of squares; V.C.- Components of Variance; P.V.- Percentage of variation

*Significant at $p < 0.001$ (999 permutation)

Table 6. Pair-by-pair matrix based on the values of genetic differentiation (F_{ST}) for the population groups of *Syagrus coronata*, where above contains the values F_{ST} from SSR data and below the F_{ST} from SNP data.

Populations	Loc. A	Loc. B	Loc. C	Loc. D	Loc. E	Loc. F
Loc. A	-	0.064	0.058	0.062	0.048	0.034
Loc. B	-0.048	-	0.030	0.025	0.089	0.057
Loc. C	0.187	0.143	-	0.039	0.093	0.072
Loc. D	0.121	0.095	-0.107	-	0.101	0.070
Loc. E	0.715	0.658	0.726	0.710	-	0.023
Loc. F	0.345	0.311	-0.052	0.014	0.755	-

Loc. A (São José da Tapera), Loc. B (Pirambú), Loc. C (Coruripe), Loc. D (Santana do Ipanema), Loc. E (Raso da Catarina) and Loc. F (Água Branca).

Table S1. Details of six populations of *Syagrus coronata* collected in the Caatinga and Atlantic Forest, indicating latitude and longitude.

Location	Sample		Longitude	Latitude	Biome
	SNP	SSR			
São José da Tapera - AL (A)	17	14	- 37.3260	- 9.5766	Caatinga
Pirambú – SE (B)	19	13	- 36.7571	- 10.6035	Atlantic Forest
Coruripe – AL (C)	16	16	- 36.0046	- 10.0030	Atlantic Forest
Santana do Ipanema – AL (D)	20	13	- 37.1540	- 9.3801	Caatinga
Raso da Catarina – BA (E)	18	19	- 38.4415	- 9.4490	Caatinga
Água Branca – AL (F)	17	19	- 37.9189	- 9.2724	Caatinga

Anexos

Normas de redação de dissertação ou tese do PPGAMGP

Dissertação constitui o produto final de pesquisas desenvolvidas em cursos de Mestrado e a Tese constitui o produto final de pesquisas desenvolvidas em cursos de Doutorado. Exigem investigações próprias à área de especialização e métodos específicos.

A Dissertação ou Tese é de responsabilidade do discente, da Comissão Orientadora e da Banca Examinadora.

Estrutura

A Dissertação ou Tese deverá ser composta de: (i) capa, (ii) páginas pré- textuais, (iii) corpo propriamente dito e, (iv) anexo (páginas pós-textuais).

A capa deverá conter a autoria, título, local e ano da aprovação. As capas encadernadas em mais de um volume deverão conter as mesmas informações acrescidas da identificação do respectivo volume. Dois (2) exemplares devem ser de capas duras de cor preta e letras amarelas.

As páginas pré-textuais serão compostas:

Primeira folha interna (página de rosto), contendo; (i) autoria, (ii) título; (iii) nota explicativa de que se trata de um trabalho de Dissertação ou Tese, mencionando o Programa de Pós-Graduação, a Universidade e o grau pretendido (Mestrado ou Doutorado); (iv) comitê de orientação e (v) local e ano de aprovação. Contará, no verso desta folha, a ficha catalográfica.

Segunda folha interna deve conter, o título, o nome do pós-graduando(a), a data de aprovação, os nomes e as assinaturas do orientador e dos participantes da Banca Examinadora, local e data.

Opcionalmente, poderão ser incluídas páginas adicionais contendo: (i) agradecimento (ii) oferecimento, (iii) dedicatória e (iv) biografia do autor, obrigatoriamente, deve conter (v) lista de símbolos, figuras, tabelas e sumário.

Folha (s) em que conste (m) o resumo em português, palavras-chave, o abstract em inglês e key words. O resumo com no máximo 800 palavras deve destacar: o local da pesquisa, delineamento estatístico, caracterização do problema, focalizar o(s) objetivo(s), síntese da metodologia, resultados obtidos e conclusões.

O corpo da Dissertação ou Tese conterá todo o trabalho impresso, avaliado e aprovado pela Banca Examinadora. O corpo poderá ser organizado na forma de capítulos.

O corpo em capítulos será composto das seções:

Capítulo I: Introdução e Referencial Teórico; Capítulos (I ou mais a depender do número de artigos científicos); e Considerações Finais (opcional). As referências bibliográficas e citações seguirão as normas da **Crop Breeding and Applied Biotechnology (CBAB)**. As referências bibliográficas deverão aparecer ao final de cada capítulo.

O anexo (páginas pós-textuais) conterá material pertinente e suplementar.

Inserir cabeçalho com citação do autor e nome da dissertação ou tese, sendo a fonte tipo arial e tamanho 10, a partir do Capítulo I até a página inicial da folha anexo(s).

Editoração

Composição tipográfica. As dissertações ou teses deverão ser impressas em forma permanente e legível, com caracteres de alta definição e de cor preta no tipo Arial tamanho 12, com espaçamento 1,5.

Notação científica e medidas. A nomenclatura científica deverá ser diferenciada contextualmente, de acordo com as normas internacionais. As unidades métricas deverão seguir o padrão do Sistema Internacional de Unidades.

Papel. Utilizar papel A-4 (210 x 297 mm) branco, e suficientemente opaco para leitura normal.

Margens. A margem esquerda deve ser de 3 cm e as outras margens de 2 cm.

Paginação. Todas as páginas textuais e pós-textuais deverão ser numeradas em sequência contínua, isto é, desde a página do Capítulo I (texto corrido), até a última página, em algarismos arábicos. A sequência deverá incluir tudo que estiver como mapas, diagramas, páginas em branco e outros. As páginas pré-textuais deverão ser numeradas, sequencialmente, como algarismos romanos minúsculos.

Ilustrações. Fotografias e outras ilustrações deverão ser montadas de forma definitiva e incluídas no corpo da Dissertação ou Tese. É admitido o uso de cores nas figuras e ilustrações. Em nenhuma circunstância dever-se-á empregar fita adesiva ou material similar para afixação de ilustrações no corpo da Dissertação ou Tese. Folhas de tamanho superior a A4 serão aceitáveis, desde que dobradas, de forma a resultar em dimensões inferiores ao tamanho do papel adotado.